

Implementasi Algoritma *K-Means Clustering* Dengan Pendekatan *Active Learning* Pada Siswa SMA Untuk Menentukan Jurusan Ke Perguruan Tinggi

Muhammad Rheza Palevi¹, Zulfahmi Indra²

¹ Ilmu Komputer, Universitas Negeri Medan, Medan, Indonesia

² Ilmu Komputer, Universitas Negeri Medan, Medan, Indonesia

Email: ¹muhammadrhezapalevi@gmail.com, ²zulfahmiindra@unimed.ac.id

Email Penulis Korespondensi: muhammadrhezapalevi@gmail.com

Article History:

Received Dec 10th, 2023

Revised Dec 28th, 2023

Accepted Jan 12th, 2024

Abstrak

Menentukan jurusan tepat yang akan diambil pada jenjang perguruan tinggi sangatlah penting diperhatikan oleh siswa SMA. Kesalahan dalam memilih jurusan akan menyebabkan siswa tidak maksimal selama berada di perguruan tinggi. Oleh karena itu, penelitian ini dilakukan untuk membantu siswa dalam menentukan jurusan ke perguruan tinggi menggunakan algoritma *K-Means Clustering* dengan penerapan *Active Learning in Machine Learning*. *K-Means Clustering* dengan penerapan *Active Learning* digunakan untuk menentukan jurusan yang tepat dengan menggunakan data kuesioner dan nilai rapor yang didapat dari pihak sekolah. Pada penelitian ini, didapat hasil bahwa 113 siswa yang menjadi responden terbagi kedalam 12 kategori atau klaster pada bidang ilmu yang berbeda-beda. Secara singkat, penerapan algoritma *K-Means Clustering* dengan pendekatan *Active Learning* menghasilkan akurasi 0,059 dan membutuhkan perbaikan terhadap data yang digunakan agar mencapai akurasi yang lebih baik

Kata Kunci : Data Mining, *K-Means Clustering*, *Active Learning in Machine Learning*, Perguruan Tinggi, Siswa SMA

Abstract

Selecting the right major for higher education is a crucial decision for high school students. An incorrect choice can lead to students not performing optimally during their college years. Therefore, this research aims to assist students in making informed decisions about their college majors by employing the K-Means Clustering algorithm with the implementation of Active Learning in Machine Learning. K-Means Clustering with Active Learning is employed to identify the suitable major by utilizing questionnaire data and report card grades provided by the school. The study results in the classification of 113 respondents into 12 different clusters corresponding to distinct academic fields. In summary, the application of the K-Means Clustering algorithm with Active Learning approach yields an accuracy of 0,059 and highlights the need for data refinement to achieve improved accuracy.

Keyword : Data Mining, *K-Means Clustering*, *Active Learning in Machine Learning*, Higher Education, High School Students

1. PENDAHULUAN

Sesuai yang berlaku di Indonesia para siswa di SMA diwajibkan memilih kejuruan yang sesuai dengan minat, bakat dan kemampuan yang dimiliki oleh setiap siswa. Pemilihan jurusan ini amat sangat berpengaruh untuk kelanjutan masa depan baik yang akan menempuh pendidikan perguruan tinggi maupun yang akan mencari pekerjaan [1]. faktor-faktor yang mempengaruhi siswa dalam pemilihan jurusan dibagi menjadi dua kelompok. Pertama, faktor dari dalam diri siswa yang meliputi minat, kepribadian, dan citra/konsep diri. Sedangkan faktor dari luar diri remaja meliputi orangtua, teman sebaya, lingkungan sosial ekonomi budaya, dan saran tes bakat minat

Dalam menentukan bidang ilmu yang cocok untuk siswa diperlukan kriteria tepat, pada umumnya dalam menentukan atau melihat keahlian dan peminatan siswa terhadap satu bidang studi adalah dengan melihat nilai akademik dan minat bakat siswa. [2]

Pihak sekolah dapat memanfaatkan perkembangan ilmu *data science*, perkembangan *data science* berkembang sangat pesat seperti halnya kebutuhan manusia. Pengolahan data berdasarkan *data science* dapat digunakan untuk mengklastering

siswa [3]. pengolahan data untuk mengklaster siswa juga dapat diterapkan pada kasus pemilihan bidang studi atau bidang ilmu yang tepat untuk siswa.

Salah satu algoritma yang digunakan untuk menentukan bidang studi siswa SMA adalah algoritma *K-Means Clustering*, Algoritma *K-means* merupakan salah satu algoritma *clustering* (pengelompokan). Hal ini juga dapat sejalan dengan penerapan *Active Learning in Machine Learning*, dalam beberapa kutipan jurnal dan literatur *Active Learning in Machine Learning* bertujuan untuk membangun secara optimal model *Supervised Learning*, optimasi model adalah salah satu cara untuk dapat menghasilkan kinerja terbaik dengan biaya yang minimum dengan klasifikasi manual.

Untuk fungsi dan kelebihan dari penerapan *Active Learning* ini sendiri adalah *Active Learning* (AL) dapat mencapai akurasi sebanding dengan *Passive Learning* dengan lebih sedikit label, hal ini sudah di amati dalam banyak pengaplikasian praktis dalam beberapa dekade terakhir. Proses dilakukan dengan memilih dan mengidentifikasi data yang paling sulit diklasifikasikan atau memiliki tingkat *incertitude* (ketidakpastian) tinggi, dan meminta annotator untuk memberikan label pada data tersebut. AL dapat digabungkan dengan algoritma *Clustering Data Mining* itu sendiri, dan juga AL tidak memiliki rumus tunggal karena AL adalah suatu pendekatan yang digunakan untuk pembelajaran mesin.

Dalam hal ini, algoritma *K-Means Clustering* dapat digunakan untuk melakukan pengelompokan siswa berdasarkan minat dan bakat mereka. Namun, hasil pengelompokan algoritma *K-Means Clustering* masih memiliki tingkat akurasi yang belum maksimal. Oleh karena itu, penerapan AL dapat menjadi solusi untuk meningkatkan akurasi hasil pengelompokan. *Active learning* meminta masukan atau validasi dari siswa untuk memvalidasi hasil clustering dan memperbaiki akurasi klasifikasi bidang studi yang diminati.

2. METODOLOGI PENELITIAN

2.1 Teknik Pengumpulan Data

Dalam menjalankan penelitian, pengumpulan data menjadi langkah penting dalam memperoleh informasi yang diperlukan. Teknik pengumpulan data yang tepat dan instrumen penelitian yang valid sangat berperan dalam menghasilkan data yang akurat dan dapat diandalkan [4]. Data yang digunakan pada penelitian ini adalah data yang bersumber dan diperoleh dari pihak sekolah SMAS Muhammadiyah 2 Medan. Data yang akan diperoleh sebanyak 113 data siswa, terdiri dari 2 jurusan. Data-data tersebut hanya akan diambil sekitar 50% dari total populasi, data akan diperoleh berupa nilai hasil kuesioner minat dan bakat siswa berupa hasil pengisian kuisisioner dan Rata-rata Nilai Rapot dari semester 1 - 4 siswa

2.2 Pre-Processing Data

Pada proses *pre-processing* data, Langkah pertama yang harus dilakukan adalah menghitung total nilai bobot disetiap rumpun ilmu dari pertanyaan angket minat/bakat, lalu untuk data nilai rapot siswa akan dihitung nilai rata-rata setiap mata pelajaran, hal ini dilakukan agar mempermudah pembuatan *dataset*, agar lebih efisien diolah oleh model maupun ketika melakukan perhitungan secara manual. Proses ini dilakukan dengan cara memanfaatkan nilai hasil kuesioner yang sudah di totalkan di setiap soalnya pada masing-masing rumpun ilmu.

2.3 Transformasi Data

Pada tahap transformasi data hasil kuesioner, peneliti menggunakan *skala likert* untuk setiap soal, peneliti akan membagi soal soal kedalam 12 kategori sesuai dengan jumlah rumpun ilmu, data yang awalnya bersifat tidak numerik yaitu kuesioner atau angket diberikan bobot nilai untuk setiap jawaban yang dipilih menggunakan metode *Skala Likert*, setiap kategori rumpun ilmu memiliki 3 soal dan setiap soal memiliki 3 pilihan jawaban sesuai dengan aturan *Skala Likert*, pilihan jawaban yang lebih memiliki kecenderungan memiliki nilai point lebih tinggi daripada 2 pilihan jawaban yang lain, sebagai contoh jika siswa memilih pilihan jawaban dengan opsi "A" maka jawaban tersebut memiliki nilai bobot 3 poin, jika "B" dengan 2 poin, dan "C" dengan 1 poin, setiap total nilai bobot pada setiap rumpun ilmu akan dijadikan variabel atau atribut data siswa.

Kajian mengenai transformasi data dari ordinal sudah pernah dilakukan oleh Muchlis dengan menggunakan bantuan macro minitab [5]. Transformasi Data adalah perubahan data dengan mengubah nilai atribut awalnya menjadi nilai atribut yang sesuai dengan kebutuhan data dalam pengolahannya [6]

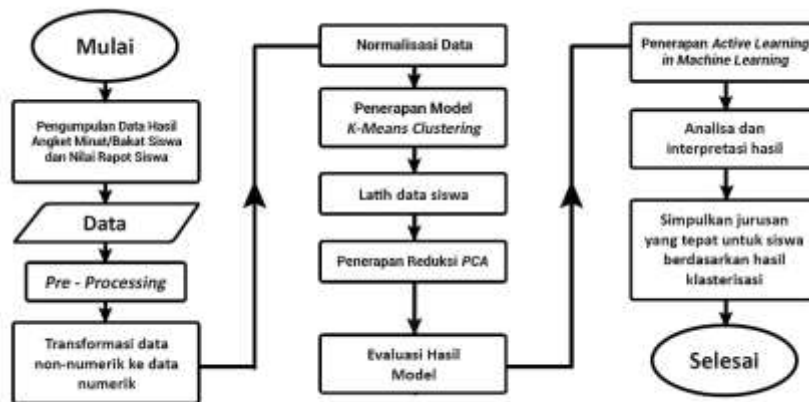
2.4 Normalisasi Data

Pada tahap berikutnya akan dilakukan tahap normalisasi data. Tujuan dari normalisasi data dalam dataset adalah untuk membentuk data dalam posisi nilai dengan rentang yang sama [7], Normalisasi data dilakukan dengan menggunakan metode *MinMaxScaler*. Tahap ini akan berguna untuk mengubah distribusi fitur-fitur dataset dengan menskalakan setiap fitur secara individual ke dalam rentang tertentu dan menghasilkan rentang nilai yang seragam pada setiap fitur, sehingga perbedaan skala antar atribut data siswa tidak mempengaruhi proses analisis secara signifikan.

2.5 Perancangan Model *K-Means Clustering*

Dengan memanfaatkan algoritma *K-Means* data yang ada akan dikelompokkan menjadi beberapa *cluster* (kelompok)

yang masing-masing diwakili oleh pusat *cluster* [8]. Pada tahap pembangunan model *K-Means Clustering* dan alur penelitian digambarkan sebagai berikut :



Gambar 1 Diagram Alir Perancangan Model *K-Means Clustering*

Pada diagram alir sebelumnya menjelaskan rangkaian langkah dalam membuat model *k-means clustering*, dimulai dengan pengumpulan data siswa (Data hasil kuesioner minat/bakat siswa dan data nilai rapot) dari SMAS Muhammadiyah 2 Medan, kemudian dilakukan tahap *pre-processing* data dan mentransformasi data yang tidak bersifat numerik menjadi data yang bersifat numerik. Pada tahap berikutnya dilakukan tahap normalisasi data pada data siswa, lalu melakukan tahap membangun model *k-means clustering* dengan menggunakan beberapa *library* yang dibutuhkan untuk proses klasterisasi dengan *library NumPy, Pandas, Matplotlib, dan Seaborn*. *Numpy* sendiri merupakan salah satu *library python* yang digunakan untuk mengimplementasi array dan matriks multidimensi [9], *Pandas* adalah *library* pada bahasa pemrograman *Python* untuk melakukan analisis data [10], *Matplotlib* adalah *library* yang bertanggung jawab untuk merencanakan data numerik [11], *Seaborn* dapat digunakan untuk membuat plot atau grafik yang menjelaskan hasil analisis [12]. Berikutnya, Model akan melakukan perhitungan, Dalam beberapa pendekatan untuk mengoptimasi *K-Means*, pemilihan *centroid* awal dilakukan dengan mengambil nilai jarak terjauh atau yang memiliki nilai densitas terbesar antar objek data [13], tentunya dengan menggunakan *Euclidean distance* yang merupakan salah satu metode perhitungan jarak yang digunakan untuk mengukur jarak dari 2 (dua) buah titik [14], lalu mendapatkan hasil klasterisasi dari *training model*, setelah melakukan perhitungan dengan *training data*, dilakukan tahap berikutnya yaitu menerapkan reduksi *Principal Component Analysis (PCA)*, *PCA* dapat mereduksi dimensi data yang tinggi menjadi dimensi data yang lebih rendah dengan resiko kehilangan informasi yang sangat kecil [15], dan melakukan perhitungan akurasi dengan menggunakan evaluasi *Silhouette Score*. Berikutnya, melakukan pendekatan *active learning* terhadap data yang bersifat tidak berlabel dan masuk kedalam kategori buruk, pada tahap setelahnya, dilakukan analisa dan interpretasi hasil untuk melihat bagaimana hasil dari model. Dari hasil evaluasi tersebut, diambil kesimpulan jurusan yang paling tepat untuk siswa berdasarkan hasil klasterisasi sebelumnya.

2.6 Pendekatan *Active Learning* Pada Model *K-Means Clustering*

Pada tahap pendekatan *active learning*, akan dilakukan beberapa tahapan untuk meningkatkan kinerja performa dari model *k-means clustering*, saat model menampilkan hasil data yang kurang baik dan tidak jelas atau sulit diklasterisasi, metode pendekatan *Active Learning* akan membantu model untuk memvalidasi hasil klasterisasi tersebut dengan bantuan *human annotator* (individu yang mendukung model untuk memberikan informasi tambahan).

2.7 Evaluasi Model *K-Means Clustering*

Pada tahap evaluasi model *k-means clustering* pada penelitian ini, akan digunakan satu metode evaluasi yang biasanya digunakan untuk model yang berbasis klasterisasi. Pada tahap ini akan mendapatkan 1 nilai sebagai parameter evaluasi, yaitu untuk melihat apakah data masuk kedalam satu klaster yang tepat atau tidak dengan metode *Silhouette Score*. Akan diterapkan sedikit *Elbow Method*, *Elbow Method* digunakan untuk optimasi algoritma *K-Means* dengan menghasilkan jumlah cluster yang optimal [16]. Sedangkan *Silhouette Score* digunakan untuk memvalidasi baik sebuah data, cluster tunggal (satu cluster dari sejumlah cluster) atau bahkan seluruh cluster [17].

3. HASIL DAN PEMBAHASAN

3.1 Collecting Data

Pada tahap pengumpulan data, peneliti melakukan beberapa tahapan dan mengumpulkan 2 jenis data. Pertama, data nilai raport siswa diperoleh melalui kerjasama dengan pihak sekolah dan sumber data kedua adalah kuesioner siswa, yang dirancang untuk menggali minat dan bakat siswa dalam berbagai bidang ilmu. Kuesioner terdiri dari 36 pertanyaan

3.1.1 Data Nilai Raport Siswa

Data nilai raport mencakup nilai-nilai siswa dari semester 1 hingga semester 4, yang terdiri dari mata pelajaran Pendidikan Agama Islam, PPKn, Bahasa Arab, Prakarya, PJOK, Seni Budaya, Bahasa Indonesia, Bahasa Inggris, Matematika, Geografi, Ekonomi, Sosiologi, Sejarah, Matematika Peminatan, Biologi, Fisika, dan Kimia.

3.1.2 Data Kuesioner Siswa

Pada data kuesioner, data dikumpulkan melalui kuesioner yang dirancang untuk mengukur minat dan bakat siswa. Kuesioner ini terdiri dari 36 pertanyaan yang menggunakan skala pilihan berganda dengan metode skala Likert. Setiap pertanyaan ditujukan untuk mengungkapkan minat dan bakat siswa dalam berbagai bidang ilmu seperti Matematika dan Ilmu Pengetahuan Alam (MIPA), Ilmu Tanaman, Ilmu Hewani, dan lainnya.

3.2 Pre-Processing

Proses perhitungan ini dilakukan untuk setiap mata pelajaran secara terpisah, menghasilkan nilai rata-rata dari seluruh nilai yang telah diperoleh oleh siswa selama masa itu.

Tabel 1. Nilai Rata-rata Siswa Pada Setiap Mata Pelajaran (Semester 1 – 4)

No	Responden (Siswa)	Jurusan	Mata Pelajaran																
			PAI	PPKn	B. Arab	Prakarya	PJOK	S. Budaya	B. Indonesia	B. Inggris	Matematika	Geografi	Ekonomi	Sosiologi	Sejarah	M. Peminatan	Biologi	Fisika	Kimia
1	R1	IPA	86	86	84	89	87	89	90	87	87	89	89	90	82	89	84	88	85
2	R2	IPA	86	91	83	88	90	84	86	86	88	87	91	93	87	87	91	91	89
3	R3	IPA	85	87	83	90	90	86	84	86	88	87	82	88	87	89	87	90	86
4	R4	IPA	84	85	87	86	92	90	85	92	84	88	84	85	86	88	88	88	85
5	R5	IPA	85	89	88	88	92	89	86	87	90	87	81	86	87	86	88	87	88
6	R6	IPA	89	84	86	92	89	89	87	91	87	86	85	88	90	87	86	89	88
7	R7	IPA	87	91	89	89	84	87	81	86	86	84	86	85	81	89	88	88	91
8	R8	IPA	88	86	88	86	91	86	90	87	86	88	85	85	88	88	87	91	88
9	R9	IPA	89	87	86	89	87	89	86	87	88	86	87	87	86	87	88	91	89
10	R10	IPA	88	83	88	88	89	89	86	88	86	85	87	88	86	85	85	87	83
...

Pada data Kuesioner, pertanyaan-pertanyaan dalam kuesioner dikelompokkan menjadi 12 bidang ilmu yang berbeda, mewakili setiap bidang minat/bakat yang diteliti.

Tabel 2. Nilai Total Kuesioner Pada Setiap Rumpun Ilmu

No	Responden (Siswa)	Jurusan	Bidang Ilmu											
			MIPA	Ilmu Tanaman	Ilmu Hewani	Ilmu Kedokteran	Ilmu Kesehatan	Ilmu Teknik	Ilmu Bahasa	Ilmu Ekonomi	Sosial Humaniora	Agama & Filsafat	Seni & Media	Ilmu Pendidikan
1	R1	IPA	7	7	8	6	7	7	8	6	6	8	7	9
2	R2	IPA	5	8	9	6	7	7	9	8	8	8	7	8
3	R3	IPA	6	7	3	7	6	6	9	9	8	6	6	8
4	R4	IPA	6	7	8	7	5	7	9	8	7	6	7	9
5	R5	IPA	7	8	8	8	7	8	8	9	8	8	8	9
6	R6	IPA	7	8	6	7	6	8	6	7	7	8	7	8
7	R7	IPA	6	8	7	5	4	7	7	7	9	6	6	6
8	R8	IPA	8	5	7	7	7	9	9	9	9	9	6	9
9	R9	IPA	7	5	9	7	6	7	5	7	8	7	7	5
10	R10	IPA	6	7	8	5	6	7	7	7	8	7	7	7
...

3.3 K-Means Clustering Dataset

Data yang digunakan dalam penelitian ini dibentuk menjadi sebuah dataset yang mencakup 29 atribut atau variabel dari data hasil kuesioner serta data mata pelajaran. Dataset ini terdiri 113 data siswa, yang merupakan total responden dalam penelitian ini. Pengolahan dataset dilakukan menggunakan aplikasi *Microsoft Excel*, di mana atribut-atribut tersebut digabungkan dalam format yang sesuai dengan persyaratan *CSV (Comma Separated Values)*, menggunakan rumus Excel yang disusun secara sistematis dan terstruktur.

Tabel 3. Set Data Siswa

MIPA	Ilmu Tanaman	Ilmu Hewani	Ilmu Kedokteran	Ilmu Kesehatan	...	M. Peminatan	Biologi	Fisika	Kimia
------	--------------	-------------	-----------------	----------------	-----	--------------	---------	--------	-------

7	7	8	6	7	...	89	84	88	85
5	8	9	6	7	...	87	91	91	89
6	7	3	7	6	...	89	87	90	86
6	7	8	7	5	...	88	88	88	85
7	8	8	8	7	...	86	88	87	88
...
6	7	8	5	8	...	88	85	87	91
6	4	6	8	9	...	87	90	87	81
6	5	8	6	9	...	84	85	84	85
4	6	6	4	6	...	86	88	86	88

3.4 Normalisasi Data

Pada penelitian ini dilakukan tahap normalisasi data. Normalisasi data dilakukan dengan menggunakan metode *MinMaxScaler*. *MinMaxScaler* digunakan untuk mengubah distribusi fitur-fitur dataset dengan menskalakan setiap fitur secara individual ke dalam rentang tertentu.

3.5 Penerapan Model *K-Means Clustering*

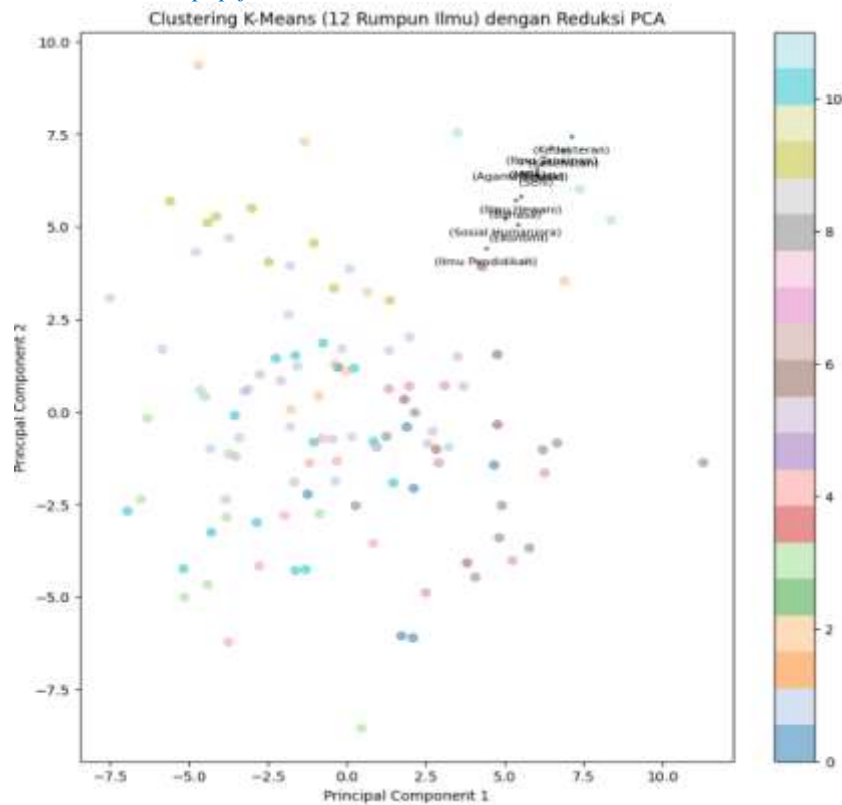
Pada tahap pembuatan model *K-Means Clustering*, Model dikonfigurasi untuk menghasilkan 12 cluster yang berbeda, yang masing-masing akan mewakili salah satu dari 12 Rumpun Ilmu yang menjadi pilihan bagi siswa. Dalam penerapan model ini, digunakan library *Scikit-Learn (Sklearn)* yang menyediakan beragam alat dan metode yang mempermudah implementasi *K-Means Clustering*. Dengan bantuan *Sklearn*, model ini dapat dibangun dengan efisien. Hasil dari model ini akan menjadi dasar penentuan bidang ilmu yang paling sesuai bagi siswa.

3.6 Hasil Proses Training

Hasil dari proses *K-Means Clustering* kemudian dicatat dan disimpan dalam format *dataset .csv* yang terstruktur. Dataset ini memuat informasi penting mengenai klasifikasi siswa ke dalam 12 cluster. Format *.csv* digunakan karena kemudahan dalam pengolahan dan kemudahan aksesibilitas data ini untuk analisis lebih lanjut.

3.7 Plotting Hasil Cluster *K-Means* Dengan Reduksi PCA

Pada tahap berikutnya yaitu melakukan *plotting cluster K-Means* dengan memanfaatkan metode reduksi *PCA*. *PCA (Principal Component Analysis)* adalah suatu teknik reduksi dimensi yang efektif dan digunakan sebagai metode analisis data. *PCA* membantu mengungkap pola tersembunyi dalam data, mengurangi kompleksitas data, dan mengekstraksi informasi penting dari dataset berdimensi tinggi.



Gambar 2 Hasil Perhitungan Model Dalam Tampilan Visual Dengan Reduksi PCA dan Anotasi Centroid

Titik berbentuk bintang kecil berwarna hitam dengan makna menggambarkan titik 12 centroid cluster yang diberi anotasi dengan nama tiap cluster yaitu cluster Kedokteran, Ilmu Tanaman, Kesehatan, MIPA, Teknik, Agama Filsafat, Seni, Ilmu Hewani, Bahasa, Sosial Humaniora, Ekonomi, dan Ilmu Pendidikan.

3.8 Hasil Clustering

Hasil perhitungan dengan menggunakan model *K-Means Clustering* telah menghasilkan pengelompokan data siswa ke dalam 12 cluster yang mewakili berbagai rumpun ilmu. Data ditampilkan dalam bentuk tabel percluster. Setiap cluster ditampilkan dengan daftar data siswa yang termasuk di dalamnya.

Tabel 4. Hasil Clustering dari Perhitungan Model *K-Means Clustering*

No	Cluster	Nama Cluster	Jurusan	Responden
1	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPA	R17
2	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPA	R20
3	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPA	R29
4	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPA	R43
5	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPA	R48
6	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPS	R83
7	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPS	R84
...
108	12	Ilmu Pendidikan	IPA	R1
109	12	Ilmu Pendidikan	IPA	R2
110	12	Ilmu Pendidikan	IPA	R25
111	12	Ilmu Pendidikan	IPA	R30
112	12	Ilmu Pendidikan	IPA	R35
113	12	Ilmu Pendidikan	IPA	R53

3.9 Evaluasi Model

Dalam upaya untuk mengukur akurasi pengelompokan data, digunakan metrik evaluasi *Silhouette Score*. *Silhouette Score* adalah metode yang mengukur sejauh mana setiap data poin sesuai dengan kelompoknya.

Gambar 3 Hasil Akurasi

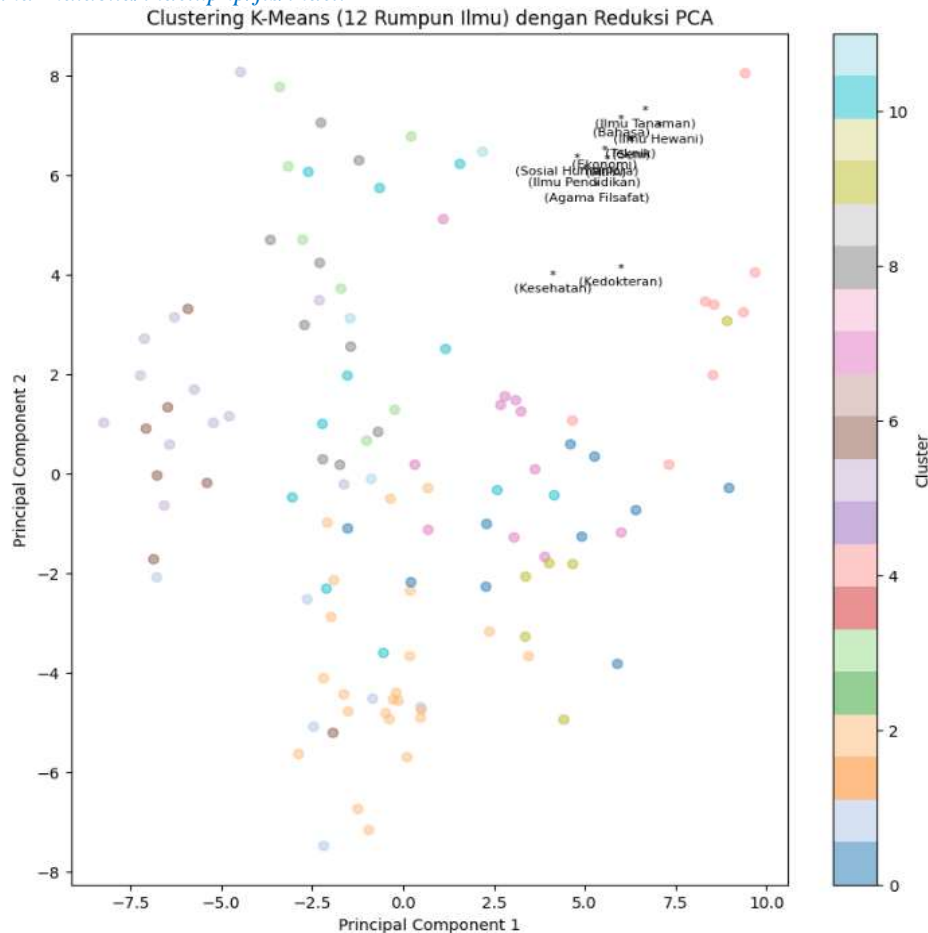
Nilai ini mengindikasikan bahwa model *K-Means Clustering* yang telah diterapkan masih memiliki ruang untuk perbaikan dalam pengelompokan data. Ini dikarenakan data primer yang diperoleh masih memiliki kekurangan untuk dihitung oleh model, akan tetapi, dari nilai akurasi yang tidak bersifat negatif dapat didefinisikan bahwa model sudah berhasil menempatkan data kedalam klaster yang tepat.

3.10 Penerapan Metode *Active Learning in Machine Learning*

Proses pendekatan *Active Learning in Machine Learning* ini melibatkan tahap penting, mulai dari proses penambahan label bantuan pada dataset hingga penghitungan ulang model. Dataset yang digunakan pada tahap ini pada awalnya dapat bersifat *unlabelled*, namun melalui pengaplikasian *Active Learning in Machine Learning*, dataset tersebut diberikan label bantuan tambahan yang dianggap lebih baik dari seorang *human annotator*.

Tabel 5. Set Data *Active Learning*

MIPA	Ilmu Tanaman	Ilmu Hewani	Ilmu Kedokteran	Ilmu Kesehatan	...	Biologi	Fisika	Kimia	Label Tambahan
7	7	8	6	7	...	84	88	85	12
5	8	9	6	7	...	91	91	89	12
6	7	3	7	6	...	87	90	86	8
6	7	8	7	5	...	88	88	85	5
7	8	8	8	7	...	88	87	88	5
...
6	7	8	5	8	...	85	87	91	12
6	4	6	8	9	...	90	87	81	2
6	5	8	6	9	...	85	84	85	12
4	6	6	4	6	...	88	86	88	1



Gambar 4 Hasil Perhitungan Model Dalam Tampilan Visual Dengan Reduksi PCA dan Anotasi Centroid Dengan Active Learning

Hasil dari pendekatan metode *Active Learning in Machine Learning* pada penelitian ini menunjukkan adanya perubahan dalam hasil *clustering*, representasi visual dengan reduksi PCA, serta akurasi model yang digunakan. Terutama, hasil klusterisasi yang awalnya menunjukkan kelompok yang agak tersebar, mengalami perubahan dan menjadi sedikit lebih terkumpul secara visual, seiring dengan metode reduksi PCA yang diterapkan.

Silhouette Score: 0.0594374193408796

Gambar 5 Hasil Akurasi Dengan Active Learning

Selain itu, akurasi model pun mengalami peningkatan dari 0,043 menjadi 0,059. Perubahan ini mengindikasikan bahwa pendekatan *Active Learning in Machine Learning* dalam penelitian ini mampu menghasilkan klusterisasi yang lebih baik dan meningkatkan kemampuan model dalam mengidentifikasi pola dalam data siswa. Hal ini memberikan keyakinan bahwa pendekatan *Active Learning in Machine Learning* dapat digunakan sebagai metode dalam meningkatkan hasil *clustering* dan akurasi dalam analisis data berbasis *cluster*.

Tabel 6. Hasil *Clustering* dari Perhitungan Model *K-Means Clustering* Dengan *Active Learning*

No	Cluster	Nama Cluster	Jurusan	Responden
1	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPA	R17
2	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPA	R20
3	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPA	R29
4	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPA	R43
5	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPA	R48

6	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPA	R56
7	1	MIPA (Matematika dan Ilmu Pengetahuan Alam)	IPS	R83
...
108	12	Seni, Media dan Desain	IPS	R70
109	12	Seni, Media dan Desain	IPS	R111
110	12	Ilmu Pendidikan	IPA	R12
111	12	Ilmu Pendidikan	IPS	R64
112	12	Ilmu Pendidikan	IPS	R80
113	12	Ilmu Pendidikan	IPS	R89

3.11 Interpretasi Hasil

Hasil akhir dari perhitungan *clustering* pada penelitian ini memberikan gambaran yang cukup komprehensif tentang minat dan bakat siswa dalam konteks bidang ilmu tertentu. Berdasarkan hasil *clustering*. Dengan merujuk pada rekomendasi jurusan yang ada dalam jurnal Kemendikbud, kita dapat mengaitkan kelompok siswa dengan jurusan-jurusan yang sesuai. Misalnya, klaster 1 (MIPA) dapat merepresentasikan siswa yang cenderung memiliki minat dalam jurusan di bidang-bidang MIPA.

4. KESIMPULAN

Penelitian ini berhasil menerapkan dan mengimplementasikan metode *K-Means Clustering* dengan pendekatan *Active Learning in Machine Learning*. Dengan menggunakan metode *K-Means Clustering* dengan pendekatan *Active Learning in Machine Learning*, pihak sekolah maupun siswa memiliki gambaran yang lebih terarah bagaimana tahapan untuk menentukan bidang ilmu yang tepat untuk siswa. Dasar referensi untuk melihat pilihan jurusan berdasarkan 12 rumpun ilmu umum dari pihak Kementerian Pendidikan dan Kebudayaan (Kemendikbud) juga sangat membantu.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

DAFTAR PUSTAKA

- [1] R. F. Putri, N. A. Ramdhan, and O. S. Bachri, "Sistem Pemilihan Jurusan Berbasis Web di SMK Wilayah Brebes," *J. Infokam*, vol. 18, no. 2, pp. 12–26, 2022.
- [2] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 690–703, 2021, doi: 10.24176/simet.v11i2.5482.
- [3] R. P. Primanda, A. Alwi, and D. Mustikasari, "DATA MINING SELEKSI SISWA BERPRESTASI UNTUK MENENTUKAN KELAS UNGGULAN MENGGUNAKAN METODE K-MEANS CLUSTERING (Studi Kasus di MTS Darul Fikri)," *Komputek*, vol. 5, no. 1, p. 88, 2021, doi: 10.24269/jkt.v5i1.686.
- [4] Ardiansyah, Risnita, and M. S. Jailani, "Teknik Pengumpulan Data Dan Instrumen Penelitian Ilmiah Pendidikan Pada Pendekatan Kualitatif dan Kuantitatif," *J. IHSAN J. Pendidik. Islam*, vol. 1, no. 2, pp. 1–9, 2023, doi: 10.61104/ihsan.v1i2.57.
- [5] S. Ningsih and H. H. Dukalang, "Penerapan Metode Suksesif Interval pada Analsis Regresi Linier Berganda," *Jambura J. Math.*, vol. 1, no. 1, pp. 43–53, 2019, doi: 10.34312/jjom.v1i1.1742.
- [6] C. Nas, "Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5," *J. Manaj. Inform.*, vol. 11, no. 2, pp. 131–145, 2021, doi: 10.34010/jamika.v11i2.5506.
- [7] Ahmad Harmain, P. Paiman, H. Kurniawan, K. Kusriani, and Dina Maulina, "Normalisasi Data Untuk Efisiensi K-Means Pada Pengelompokan Wilayah Berpotensi Kebakaran Hutan Dan Lahan Berdasarkan Sebaran Titik Panas," *Tek. Teknol. Inf. dan Multimed.*, vol. 2, no. 2, pp. 83–89, 2022, doi: 10.46764/teknimedia.v2i2.49.
- [8] M. S. Said and Y. Yusti, "Penerapan Algoritma K-Means Dalam Penentuan Jurusan Siswa Sman 05 Bombana," *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 5, no. 2, pp. 114–122, 2020, doi: 10.51876/simtek.v5i2.87.
- [9] A. M. Chalik, B. A. Qowy, F. Hanafi, and ..., "Mouse Tracking Tangan dengan Klasifikasi Gestur Menggunakan OpenCV dan Mediapipe," ... *Ilm. Tek. Inform.* ..., vol. 1, no. 2, pp. 10–18, 2021, [Online]. Available: <http://journal.sinov.id/index.php/juitik/article/view/323>
- [10] Y. Christianto, R. Intan, and R. Adipranata, "Penerapan Metode Klasifikasi C4. 5 dalam Pembuatan Website Identifikasi untuk Prediksi Kredibilitas Akun pada Media Sosial Instagram," *J. Infra*, 2021, [Online]. Available: <http://publication.petra.ac.id/index.php/teknik-informatika/article/view/11443>
- [11] I. Luthfiah Ramadhyagita *et al.*, "SPEKTRA: Jurnal Fisika dan Aplikasinya Kajian Discrete Fourier Transform untuk Menganalisis Sinyal Arbitrer," *Kaji. Discret. Fourier Transform untuk Menganalisis Sinyal Arbitrer*, vol.

- 1, pp. 7–16, 2022.
- [12] E. Sera, “Analisis Sentimen Ulasan Produk di E-Commerce Bukalapak Menggunakan Natural Language Processing,” *J. Pros. Semin. Nas. Sist. Inf. dan Teknol.*, vol. 7, pp. 237–243, 2023.
- [13] H. B. Pramesti, “Klasterisasi Data Unsupervised Menggunakan Metode K-Means (Studi Kasus: Repository Jurusan Ilmu Komputer FMIPA Universitas Riau),” *J. Apl. Komput.*, pp. 33–45, 2022, [Online]. Available: <https://jurnalmipa.unri.ac.id/jak/index.php/JAK/article/view/33>
- [14] M. Nishom, “Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square,” *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 20–24, 2019, doi: 10.30591/jpit.v4i1.1253.
- [15] A. S. Ritonga and I. Muhandhis, “Teknik Data Mining Untuk Mengklasifikasikan Data Ulasan Destinasi Wisata Menggunakan Reduksi Data Principal Component Analysis (Pca),” *EduTic - Sci. J. Informatics Educ.*, vol. 7, no. 2, 2021, doi: 10.21107/edutic.v7i2.9247.
- [16] F. Sutomo *et al.*, “Optimization of the K-Nearest Neighbors Algorithm Using the Elbow Method on Stroke Prediction,” *J. Tek. Inform.*, vol. 4, no. 1, pp. 125–130, 2023, doi: 10.52436/1.jutif.2023.4.1.839.
- [17] D. J. Lubis and M. B. Tamam, “Penerapan K-Means Untuk Pengelompokkan Beasiswa Santri di Pondok Pesantren Miftahul Huda Bogor,” *Teknois J. Ilm. Teknol. Inf. dan Sains*, vol. 12, no. 1, pp. 7–20, 2022, doi: 10.36350/jbs.v12i1.125.