

IMPLEMENTASI *CLUSTERING* UNTUK MENENTUKAN EFEKTIVITAS NILAI SISWA SESUDAH PANDEMI COVID-19 MENGGUNAKAN ALGORITMA *K-MEANS*

Hizkia Alexander, Yuyun Umaidah, M. Jajuli

Program Studi Informatika S1, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
Jl. HS. Ronggo Waluyo, Telukjambe Timur, Karawang, Indonesia
1910631170084@student.unsika.ac.id

ABSTRAK

SMKN 1 Jakarta adalah sebuah lembaga pendidikan kejuruan negeri yang berlokasi di Jakarta. Sebagai salah satu institusi pendidikan tertua di Jakarta, sekolah ini menawarkan 11 bidang keahlian atau jurusan kepada siswa. Meskipun awalnya proses pembelajaran berjalan secara normal, situasinya berubah pada awal tahun 2020 ketika muncul wabah penyakit baru yang dikenal dengan nama COVID-19. Dampak penyebaran virus yang luas mengakibatkan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (KEMDIKBUD) merilis surat edaran untuk mengadopsi pembelajaran jarak jauh atau daring. Hal ini menyebabkan variasi nilai siswa karena tantangan adaptasi terhadap metode pembelajaran yang baru. Untuk mengatasi permasalahan tersebut, peneliti menerapkan metode CRISP-DM dengan fokus pada analisis *cluster* atau pengelompokan menggunakan algoritma *K-Means* terhadap nilai siswa dalam menilai efektivitas perubahan metode pembelajaran pada semester ganjil dan genap tahun pelajaran 2021/2022. Hasil dari proses pengelompokan teridentifikasi *cluster* 0 dikategorikan sebagai kelompok dengan nilai yang baik, sementara *cluster* 1 memiliki nilai yang kurang baik. Analisis lebih lanjut menunjukkan bahwa *cluster* 2 memiliki perbedaan tertinggi menurut *Elbow Method*, dengan selisih nilai sebesar 79963.30076 pada semester ganjil, dan 212024.9629 pada semester genap. Penilaian berdasarkan *Silhouette Coefficient* juga mempertegas bahwa *cluster* 2 merupakan kelompok paling efektif dengan skor tertinggi, yaitu 0.8171218916694154 pada semester ganjil dan 0.8932255100239208 pada semester genap. Sehingga disimpulkan hasil pengelompokan menunjukkan peningkatan efektivitas sebesar 2.5%, yang mengindikasikan perbaikan nilai pada 5 siswa.

Kata kunci: *Clustering, Algoritma K-Means, Elbow Method, Silhouette Coefficient.*

1. PENDAHULUAN

Sebelum COVID-19, pendidikan adalah proses tatap muka di sekolah dengan interaksi sosial, fasilitas sekolah lengkap, dan evaluasi dalam kelas. Setelah COVID-19, pendidikan mengalami perubahan drastis, beralih ke pembelajaran jarak jauh dengan penggunaan teknologi, mengakibatkan tantangan seperti interaksi terbatas, perubahan pola belajar, dan evaluasi yang rumit. Meskipun demikian, adaptasi cepat dan inovasi dalam pengajaran terjadi, mendorong inklusi digital dan penekanan pada kualitas pembelajaran dalam situasi yang berubah-ubah. SMKN 1 Jakarta merupakan sekolah menengah kejuruan negeri yang berada di Jakarta. SMKN 1 Jakarta memiliki total 11 kompetensi keahlian atau jurusan yang tersedia. Awalnya pembelajaran di SMKN 1 Jakarta berlangsung normal namun hingga terjadinya pandemi COVID-19 dan dikeluarkannya surat yang berasal dari KEMDIKBUD pada bulan Maret 2020 terpaksa SMKN 1 Jakarta harus melakukan proses pengajaran secara daring [1].

COVID-19 adalah salah satu jenis penyakit baru yang termasuk ke dalam kelompok virus Corona, penyakit ini disebabkan oleh virus yang bernama SARS-2 CoV-2. Penyakit ini ditemukan pada akhir tahun 2019 atau lebih tepatnya pada bulan Desember di Wuhan, Provinsi Hubei, Tiongkok [2]. Penyebaran COVID-19 umumnya terbagi menjadi dua, melalui air

liur orang yang terinfeksi dan barang atau permukaan yang terkontaminasi.

Menurut *website* resmi KEMENKES RI perihal infeksi COVID-19 pada Indonesia, Jakarta menjadi provinsi dengan total kasus terinfeksi terbanyak, disusul dengan Jawa Barat sebanyak 1,250,314 total kasus, hal tersebut disampaikan pada Gambar 1.



Gambar 1. *Dashboard COVID-19 di Indonesia*
(Sumber: [kemkes](https://kemkes.go.id))

Berdasarkan permasalahan diatas, peneliti bertujuan untuk membandingkan prestasi belajar siswa sebelum dan sesudah masa pandemi guna menilai efektivitas pembelajaran daring dibandingkan metode konvensional. Walaupun pada tanggal 5 Mei 2023 COVID-19 telah dinyatakan bukan lagi sebagai darurat kesehatan global [3] namun meskipun

demikian pada konferensi pers pihak KEMENKES menjawab masih akan mempertimbangkan soal pencabutan status kedaruratan COVID-19 di Indonesia [4]. Untuk mendapatkan pemahaman yang mendalam, peneliti mengusulkan penerapan metode *clustering* dengan menggunakan algoritma *k-means*. Pendekatan ini diharapkan dapat mengidentifikasi pola dan perbedaan dalam hasil pembelajaran, yang pada akhirnya akan memberikan wawasan penting mengenai keberhasilan masing-masing pendekatan pembelajaran.

Clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum [5].

K-Means adalah salah satu algoritma *clustering* bertipe *centroid based algorithm*, *K-Means* sendiri merupakan algoritma iteratif yang mengelompokkan data menjadi *k* kelompok, pada awalnya pusat kelompok awal ditempatkan secara acak dan kemudian titik data ditetapkan ke kelompok berdasarkan jarak *euclidean* terdekat ke pusat kelompok, proses ini berlanjut hingga konvergensi, di mana posisi pusat kelompok diperbarui untuk mencapai solusi yang optimal [6].

Demi memastikan bahwa proyek *data mining* dilakukan dengan langkah-langkah yang sistematis dan menghasilkan hasil capaian yang bermakna sesuai dengan tujuan. Peneliti menggunakan CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai metodologinya, karenanya pendekatan yang digunakan dalam pengembangan proyek *data mining* ini membantu analisa data untuk merencanakan, melaksanakan, dan mengevaluasi proyek-proyek *data mining* secara sistematis [7].

2. TINJAUAN PUSTAKA

2.1. Efektivitas

Efektivitas adalah kata lain dari keefektifan, berasal dari kata efektif yang memiliki empat arti. Pertama ada efeknya, arti yang kedua adalah manjur atau mujarab, arti yang ketiga adalah dapat membawa hasil, dan arti yang terakhir adalah mulai berlaku [8]. Jadi dengan kata lain efektivitas bisa dikatakan sebagai efek atau hasil yang dihasilkan setelah melakukan sesuatu.

2.2. COVID-19

COVID-19 adalah salah satu jenis penyakit baru yang termasuk ke dalam kelompok virus *corona*, penyakit ini disebabkan oleh virus yang bernama SARS-2 CoV-2. Penyakit ini ditemukan pada akhir tahun 2019 atau lebih tepatnya pada bulan Desember di Wuhan, Provinsi Hubei, Tiongkok [2].

2.3. Data Mining

Data mining adalah proses mengekstraksi pola menarik, pengetahuan, atau informasi yang berguna dari sekumpulan besar data dengan menggunakan

teknik statistik, kecerdasan buatan, *machine learning*, dan metode lainnya [6]. *Data mining* secara umum dibagi menjadi dua kategori, yaitu *descriptive* yang bertujuan untuk mendeskripsikan data yang ada di dalam *dataset* dan *predictive* yang bertujuan untuk memahami data pada saat ini demi melakukan prediksi.

2.4. CRISP-DM

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah suatu proses standar yang terdiri dari enam tahapan utama, yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyebaran. [9] Terdapat 6 tahapan CRISP-DM beserta pengertiannya sebagai berikut:

1. *Business Understanding*
Pemahaman tentang tujuan proyek yang akan dilakukan terutama dari perspektif bisnis.
2. *Data Understanding*
Pemahaman data dengan melakukan empat langkah, yaitu: mengambil data awal, mendeskripsikan data, mengeksplorasi data, memverifikasi kualitas data.
3. *Data Preparation*
Pada tahap ini melakukan persiapan data dengan melakukan lima langkah, yaitu: menyeleksi data, membersihkan data, membangun data, mengintegrasikan data, dan pemformatan data.
4. *Modeling*
Pada tahap ini melakukan pengembangan dan pemodelan *data mining* dengan melakukan empat langkah, yaitu: *the selection of the modeling technique, the generation of test design, the creation of models, and the assessment of models.*
5. *Evaluation*
Pada tahap ini model yang telah dikembangkan dievaluasi terlebih dahulu untuk memastikan bahwa model tersebut sesuai atau memenuhi kriteria bisnis
6. *Deployment*
Mengimplementasi model yang telah dibuat.

2.5. Python

Python adalah bahasa pemrograman tingkat tinggi yang bersifat *open source*, bahasa pemrograman ini diciptakan oleh Guido van Rossum, dan bahasa pemrograman sering digunakan banyak orang terutama pada bidang data, *Artificial Intelligence* (AI), *Machine Learning* (ML) dikarenakan bahasanya yang fleksibel dan mudah dimengerti.

2.6. Google Colaboratory

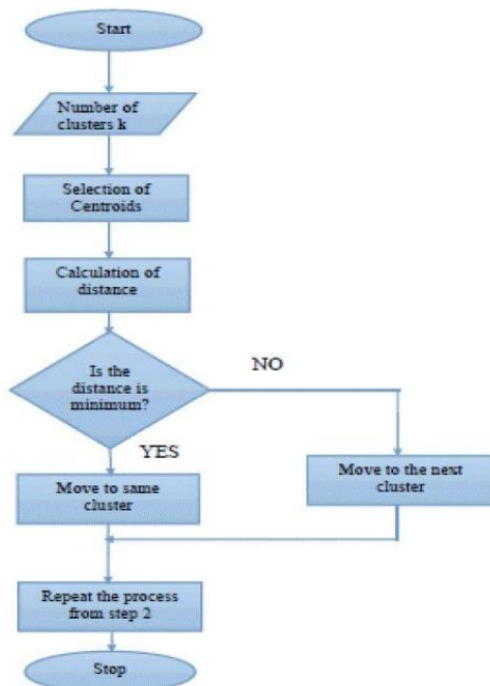
Google Colaboratory atau yang biasa disingkat *Google Colab* adalah platform pengembangan dan eksekusi kode *Python* berbasis *cloud* yang diciptakan oleh *Google*. *Google Colab* memungkinkan pengguna untuk menulis, menjalankan, dan berbagi kode *Python* secara *online* melalui peramban web.

2.7. Clustering

Clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum [5]. Meskipun *clustering* termasuk *unsupervised learning* (tanpa diarahkan atau tanpa label), *clustering* memanfaatkan karakter atau pola pada data yang mirip atau serupa. Terdapat berbagai macam tipe dari *clustering*, contohnya: *density based clustering*, *centroid based clustering*, *hierarchical based clustering*, *distribution based clustering*.

2.8. K-Means

K-Means merupakan algoritma iteratif yang mengelompokkan data menjadi k kelompok, pada awalnya pusat kelompok awal ditempatkan secara acak dan kemudian titik data ditetapkan ke kelompok berdasarkan jarak *euclidean* terdekat ke pusat kelompok, proses ini berlanjut hingga konvergensi, di mana posisi pusat kelompok diperbarui untuk mencapai solusi yang optimal [6]. *K-Means* merupakan salah satu algoritma *clustering* bertipe *centroid based*. Berikut *flowchart* dalam algoritma *K-Means* yang dijelaskan pada Gambar 2.



Gambar 2. *Flowchart* langkah-langkah *K-Means* (Sumber: modul *clustering* Ruangguru)

Untuk mendapatkan pemahaman yang lebih terperinci mengenai Gambar 2, informasinya akan diuraikan sebagai berikut:

1. Menentukan jumlah nilai K *cluster*.
2. Acak pusat *centroid* sebanyak K *cluster*.
3. Melakukan penghitungan *cluster*:

- a. Menghitung *euclidean distance* dari setiap sampel ke masing-masing *centroid*. Sampel dengan jarak terdekat ke *centroid* tertentu akan masuk ke dalam *cluster* yang sama.
 - b. Untuk setiap K , hitung rata-rata dari seluruh anggota *cluster*. Rata-rata tersebut akan menjadi *centroid* baru dalam *cluster* tersebut.
4. Jika sudah mencapai batas iterasi maksimum dan jika nilai *centroid* tidak terlalu berubah atau berpindah yang ditentukan oleh suatu batas toleransi perubahan maka proses akan berhenti.

2.9. Elbow Method

Elbow Method merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan menggunakan SSE (*Sum of Squared Errors*) antara jumlah *cluster* yang akan membentuk siku pada suatu titik [10].

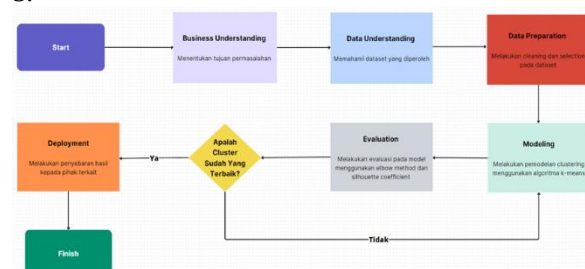
2.10. Silhouette Coefficient

Metrik ini menggabungkan tingkat kohesi (seberapa dekat titik data dengan anggota *cluster* lainnya) dan tingkat pemisahan (seberapa jauh titik data dari anggota *cluster* lainnya) dalam satu nilai. *Silhouette Coefficient* (SC) berkisar dari -1 hingga 1, dengan nilai positif menunjukkan bahwa titik data cocok dalam *cluster* mereka dan nilai negatif menunjukkan ketidakcocokan. Berikut ini adalah interpretasi nilai *Silhouette Coefficient* [11]:

1. *High SC value* (mendekati angka 1): Sampel terpisah dengan baik dengan *cluster* disekitarnya atau yang berarti sampel terkluster dengan tepat.
2. *SC value* sekitar angka 0: Sampel terkluster masih belum tepat atau dikatakan masih belum signifikan dikarenakan jarak antar sampel dengan *cluster* sekitarnya.
3. *Low SC value* (mendekati angka -1): Sampel terkluster dengan tidak tepat atau sampel ditempatkan pada *cluster* yang salah..

3. METODE PENELITIAN

Pada penelitian ini, metodologi yang digunakan adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri dari enam tahap dalam pengolahan data, hal tersebut bisa dilihat pada Gambar 3.



Gambar 3. *Flow* CRISP-DM (Sumber: [researchgate](https://researchgate.net/publication/312111111))

Pada Gambar 3, dijelaskan bahwa langkah pertama adalah *Business Understanding* di mana fokusnya adalah menetapkan tujuan penelitian serta kebutuhan yang diperlukan. Kedua, *Data Understanding* digunakan untuk memahami kualitas dan identifikasi data yang tersedia. Ketiga, *Data Preparation* melibatkan penataan data dengan menghapus variabel yang tidak relevan. Keempat, *Modeling* menggunakan metode *K-means* untuk klusterisasi dengan penerapan dua uji algoritma, yaitu *Elbow Method* dan *Silhouette Coefficient*. Kelima, *Evaluation* hasil klusterisasi, jika hasilnya kurang memuaskan, revisi model diperlukan hingga diperoleh hasil yang memadai. Keenam *Deployment*, di mana hasil dari penelitian bisa diaplikasikan atau dipublikasikan.

4. HASIL DAN PEMBAHASAN

Pada tahap *business understanding*, peneliti menggunakan empat tahap dalam pengerjaannya. Pertama ada *Determine Business Objective*, dimana penelitian berfokus pada analisis *cluster* untuk mengukur kemajuan siswa SMKN 1 Jakarta pasca COVID-19, dengan tujuan meningkatkan metode pembelajaran. Langkah kedua ada *Assess Situation*, melibatkan evaluasi ketersediaan data yang relevan. Ketiga ada *Determine Data Mining Goals*, yang dimana keberhasilan akan tercapai jika analisis *cluster* mengidentifikasi jumlah *cluster* optimal dalam analisis *cluster* untuk mengukur efektivitas. Terakhir, langkah keempat *Plan Activities*, Agar tujuan dapat tercapai dibutuhkan *tools* berupa *Google Colaboratory* dengan bahasa pemrograman *Python*.

Pada tahap *data understanding*, *dataset* yang diperoleh terdiri dari 14 variabel nilai mata pelajaran dan variabel absensi dari total 209 siswa pada semester 1 dan 2, tahun ajaran 2021-2022. Data tersebut kemudian diunggah ke *Google Drive* untuk dihubungkan langsung dengan *Google Colaboratory*. Hasil *import dataset* tersebut ditampilkan pada Gambar 4 dan Gambar 5.

	NAMA	Agama dan Budi Pekerti	PPKN	Bahasa Indonesia	Matematika (Umum)	Sejarah Indonesia	Bahasa Inggris	Seni Budaya	PJOK	Simulasi dan Komunikasi Digital	Fisika	Kimia	Mapel Jurusan 1	Mapel Jurusan 2	Mapel Jurusan 3	Absensi
8	ALTA FAKHRIDHAR	75	77	75	82	80	78	79	82	75	81	81	75	82	84	0
1	ANGGA PUTRA RAMADHAN	75	77	75	77	79	75	76	79	75	79	75	75	81	79	0
2	ARREVO FARIS NAFWAN	81	77	75	79	79	85	83	81	75	78	78	75	81	84	0
3	BUDI SATRIO	78	79	78	76	84	75	80	85	75	78	81	75	77	77	0
4	CINDE RIVA RUTHA	77	77	76	84	83	75	77	82	75	78	78	75	81	82	3

Gambar 4. Contoh *Dataset* Semester Ganjil

	NAMA	Agama dan Budi Pekerti	PPKN	Bahasa Indonesia	Matematika (Umum)	Sejarah Indonesia	Bahasa Inggris	Seni Budaya	PJOK	Simulasi dan Komunikasi Digital	Fisika	Kimia	Mapel Jurusan 1	Mapel Jurusan 2	Mapel Jurusan 3	Absensi
8	ALTA FAKHRIDHAR	77	78	81	79	82	80	80	77	77	79	80	77	79	81	1
1	ANGGA PUTRA RAMADHAN	79	80	79	76	82	78	84	83	77	79	77	77	79	81	4
2	ARREVO FARIS NAFWAN	81	81	79	77	81	82	76	83	77	79	78	77	81	78	0
3	BUDI SATRIO	77	81	78	76	81	78	83	80	77	78	79	77	81	79	0
4	CINDE RIVA RUTHA	79	80	80	80	80	80	84	84	77	79	80	77	79	81	0

Gambar 5. Contoh *Dataset* Semester Genap

Setelah itu dilakukan pemeriksaan tipe data pada kedua kolom *dataset* tersebut ditunjukkan pada Gambar 6.

df_ganjil.dtypes	df_genap.dtypes
NAMA	object
Agama dan Budi Pekerti	int64
PPKN	int64
Bahasa Indonesia	int64
Matematika (Umum)	int64
Sejarah Indonesia	int64
Bahasa Inggris	int64
Seni Budaya	int64
PJOK	int64
Simulasi dan Komunikasi Digital	int64
Fisika	int64
Kimia	int64
Mapel Jurusan 1	int64
Mapel Jurusan 2	int64
Mapel Jurusan 3	int64
Absensi	int64
dtype: object	dtype: object

Gambar 6. Tipe Data Kedua *Dataset*

Dalam Gambar 6, terlihat setiap kolom pada kedua *dataset* tersebut sudah sesuai dengan apa yang diinginkan, lalu dilanjutkan melakukan kalkulasi statistik pada kedua *dataset* tersebut untuk menemukan total, persebaran dan statistik nilai data yang ditunjukkan pada Gambar 7 dan Gambar 8.

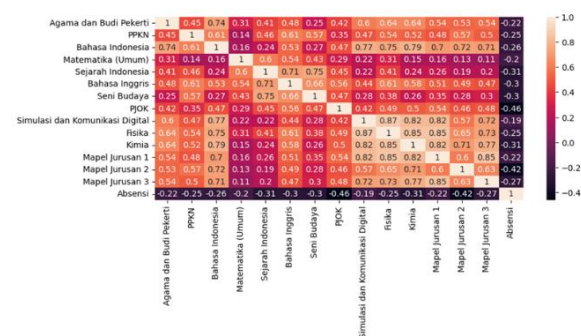
	Agama dan Budi Pekerti	PPKN	Bahasa Indonesia	Matematika (Umum)	Sejarah Indonesia	Bahasa Inggris	Seni Budaya	PJOK	Simulasi dan Komunikasi Digital	Fisika	Kimia	Mapel Jurusan 1	Mapel Jurusan 2	Mapel Jurusan 3	Absensi
count	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000
mean	75.384615	76.657143	81.342857	78.342857	78.342857	81.342857	81.342857	81.342857	81.342857	78.342857	78.342857	78.342857	81.342857	81.342857	81.342857
std	7.861916	8.943596	7.732258	11.588465	10.320587	8.123678	8.123678	8.123678	8.123678	8.123678	8.123678	8.123678	8.123678	8.123678	8.123678
min	25.000000	0.000000	0.000000	0.000000	0.000000	38.000000	2.000000	73.000000	0.000000	0.000000	0.000000	0.000000	0.000000	4.000000	0.000000
25%	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	0.000000
50%	80.000000	80.000000	82.000000	80.000000	82.000000	80.000000	80.000000	80.000000	80.000000	80.000000	80.000000	80.000000	80.000000	80.000000	0.000000
75%	82.000000	81.000000	80.000000	81.000000	80.000000	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	0.000000
max	87.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	32.000000

Gambar 7. *Describe Dataset* Ganjil

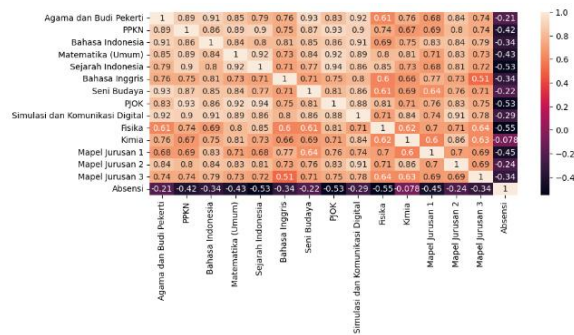
	Agama dan Budi Pekerti	PPKN	Bahasa Indonesia	Matematika (Umum)	Sejarah Indonesia	Bahasa Inggris	Seni Budaya	PJOK	Simulasi dan Komunikasi Digital	Fisika	Kimia	Mapel Jurusan 1	Mapel Jurusan 2	Mapel Jurusan 3	Absensi
count	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000	209.000000
mean	75.384615	76.657143	81.342857	78.342857	78.342857	81.342857	81.342857	81.342857	81.342857	78.342857	78.342857	78.342857	81.342857	81.342857	81.342857
std	7.861916	8.943596	7.732258	11.588465	10.320587	8.123678	8.123678	8.123678	8.123678	8.123678	8.123678	8.123678	8.123678	8.123678	8.123678
min	25.000000	0.000000	0.000000	0.000000	0.000000	38.000000	2.000000	73.000000	0.000000	0.000000	0.000000	0.000000	0.000000	4.000000	0.000000
25%	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	0.000000
50%	80.000000	80.000000	82.000000	80.000000	82.000000	80.000000	80.000000	80.000000	80.000000	80.000000	80.000000	80.000000	80.000000	80.000000	0.000000
75%	82.000000	81.000000	80.000000	81.000000	80.000000	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	0.000000
max	87.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	86.000000	32.000000

Gambar 8. *Describe Dataset* Genap

Pada Gambar 7 dan Gambar 8, ditemukan bahwa data tampak normal dari persebaran, nilai maupun jumlah datanya. Setelah itu dilakukan proses untuk menampilkan korelasi antar kolom pada kedua *dataset* tersebut yang ditunjukkan pada Gambar 9 dan Gambar 10, dan diketahui pada Gambar 10 diperoleh peningkatan korelasi antar kolomnya.



Gambar 9. Korelasi *Dataset* Ganjil



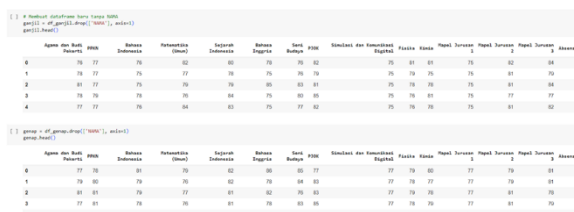
Gambar 10. Korelasi *Dataset* Genap

Pada tahap selanjutnya yaitu *data preparation*, ditahap ini dilakukan proses pembersihan, pemilihan, dan transformasi data agar data siap untuk dilakukan pemodelan. Pada Gambar 11. dilakukan proses pengecekan data dan diketahui kedua *dataset* tersebut sudah bersih dari nilai *null*.



Gambar 11. Jumlah *Null* Kedua *Dataset*

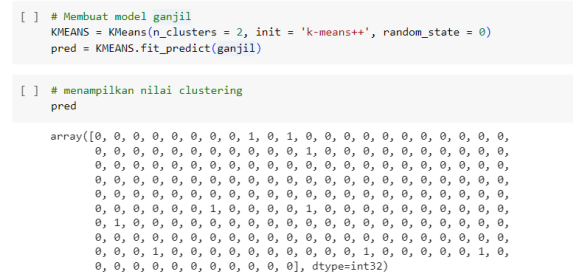
Selanjutnya yaitu melakukan proses pemilihan kolom yang akan ditampilkan pada Gambar 12.



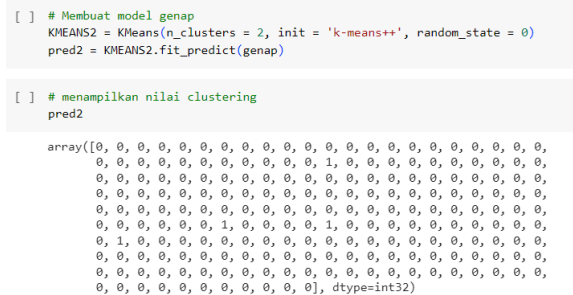
Gambar 12. Pemilihan Kolom Kedua *Dataset*

Pada Gambar 12. dilakukan pemilihan kolom yang dimana hanya kolom “NAMA” yang akan dihapus.

Masuk ke tahap *modeling*, hal pertama yang perlu dilakukan adalah menentukan jumlah *cluster*; disini peneliti akan menggunakan dua *clusters* terlebih dahulu untuk digunakan pada *dataset* semester ganjil dan genap, yang selanjutnya *cluster* akan dievaluasi menggunakan *Elbow Method* dan *Silhouette Coefficient*. Setelah hasil prediksi *cluster* didapatkan selanjutnya hasil akan ditampilkan masih berupa array pada Gambar 13 dan Gambar 14.



Gambar 13. Prediksi *Cluster* Ganjil

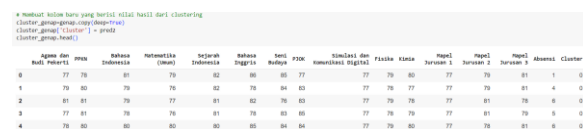


Gambar 14. Prediksi *Cluster* Genap

Setelah itu kolom hasil prediksi akan ditampilkan contohnya bersama kolom mata pelajaran, Absensi dan NAMA yang sudah di buang sebelumnya

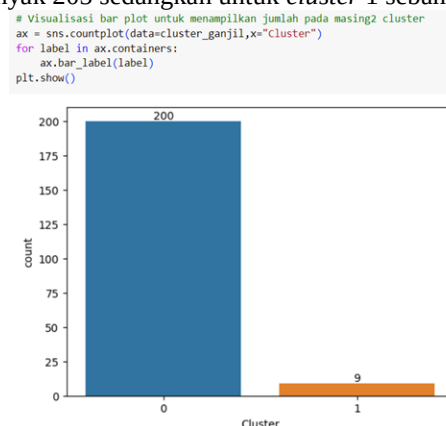


Gambar 15. Hasil *Cluster* Ganjil



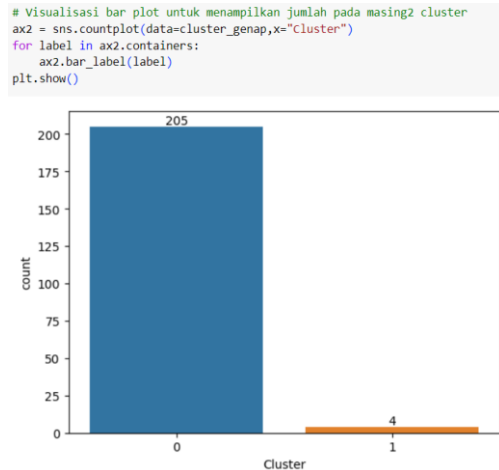
Gambar 16. Hasil *Cluster* Ganjil

Kemudian dilakukan visualisasi *cluster* semester ganjil yang ditunjukkan pada Gambar 4.14 digambarkan *cluster* 0 berjumlah sebanyak 200 sedangkan untuk *cluster* 1 sebanyak 9. Untuk visualisasi *cluster* genap yang ditunjukkan pada Gambar 4.15 digambarkan *cluster* 0 berjumlah sebanyak 205 sedangkan untuk *cluster* 1 sebanyak 4.



Cluster

Gambar 17. Visualisasi *Cluster* Ganiil



Gambar 18. Visualisasi Cluster Genap

Dari Gambar 4.14 dan Gambar 4.15 diperoleh terjadi perubahan bertambahnya nilai cluster 0 sebanyak 5. Selanjutnya pada Tabel 4.1 akan ditampilkan contoh siswa yang termasuk cluster 0 dan 1 pada semester ganjil dan pada Tabel 4.2 akan ditampilkan contoh siswa yang termasuk cluster 0 dan 1 pada semester genap.

Tabel 1. Hasil Clustering Ganjil

Nama	Agama dan Budi Pekerti	PPKN	...	Absensi	Cluster
ALYA FAAKHIRAH	76	77	...	0	0
ANDIKA PUTRA RAMADHAN	78	77	...	6	0
ARIEYO FARISI NAFHAN	81	77	...	6	0
BUDI SATRIO	78	79	...	6	0
CINDE REVA MUTIA	77	77	...	3	0
FADILAH NUR ABIA	32	76	...	8	1
HAWA AZ ZAHRA	75	0	...	10	1
WINDY PRATAMA	75	75	...	8	1
FARREL ARDAN REVALINO	25	0	...	8	1
HIKMAL ERLANGGA GUNAWAN	31	76	...	0	1
RAFID NAUFAL SANJAYA	76	78	...	5	1
BAGAS ADE PRASETYO	75	64	...	6	1
MUAMAR KADAFI	75	76	...	6	1
MUHAMMAD RIFKI	77	76	...	6	1

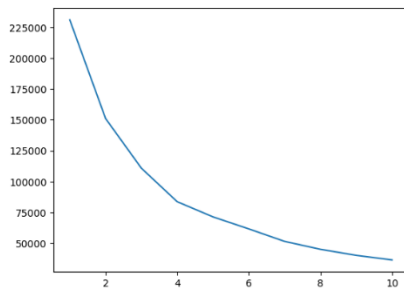
Tabel 2. Hasil Clustering Genap

Nama	Agama dan Budi Pekerti	PPKN	...	Absensi	Cluster
ALYA FAAKHIRAH	77	78	...	1	0
ANDIKA PUTRA RAMADHAN	79	80	...	4	0
ARIEYO FARISI NAFHAN	81	81	...	6	0
BUDI SATRIO	77	81	...	5	0
CINDE REVA MUTIA	78	80	...	6	0
WINDY PRATAMA	40	30	...	20	1
FARREL ARDAN REVALINO	0	0	...	4	1
HIKMAL ERLANGGA GUNAWAN	0	0	...	0	1
RAFID NAUFAL SANJAYA	0	0	...	4	1

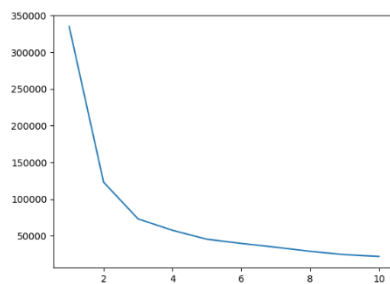
Dari Tabel 1 dan Tabel 2. terlihat bahwa cluster 0 bisa dinyatakan sebagai cluster dengan tingkat nilai yang baik sedangkan pada cluster 1 terlihat dengan tingkat nilai yang kurang baik. Selain itu terdapat penambahan berjumlah 5 orang pada cluster 0 yang dengan kata lain terjadi peningkatan nilai siswa sebanyak 5 orang pada semester genap.

Setelah dilakukan pemodelan hal yang perlu dilakukan adalah evaluasi, gunanya untuk mengetahui tingkat keberhasilan suatu pemodelan. Pada tahap evaluasi ini peneliti akan menggunakan metode *Elbow Method* dan *Silhouette Coefficient* pada kedua hasil pemodelan cluster yang sudah dilakukan. Pada Gambar 19 dan Gambar 20. telah dilakukan evaluasi *Elbow Method* pada hasil pemodelan semester ganjil

dan genap, dari gambar tersebut diperoleh grafik nilai *cluster* 1-10.



Gambar 19. *Elbow Method* Ganjil



Gambar 20. *Elbow Method* Genap

manakah yang nilai setelahnya turun atau naik dengan stabil seterusnya. Seperti yang terlihat pada Gambar 19 dan Gambar 20, nilai tampak stabil di antara *cluster* 2 atau 3. Berikut data detailnya dari Gambar 19 dan Gambar 20, yang disampaikan pada Tabel 3 dan Tabel 4.

Tabel 3. Nilai *Elbow Method* Ganjil

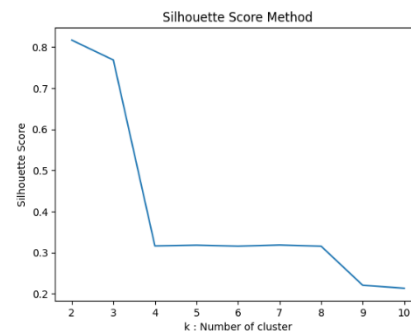
Cluster	Nilai	Selisih
1	231028.57416267938	
2	151065.2733990148	79963.30076
3	110800.44006568144	40264.83333
4	83655.75053402505	27144.68953
5	71357.58386735835	12298.16667
6	61603.513073209964	9754.070794
7	51489.917200691685	10113.59587
8	45068.02064896753	6421.896552
9	40170.581684981684	4897.438964
10	36463.08168498169	3707.5

Tabel 4. Nilai *Elbow Method* Genap

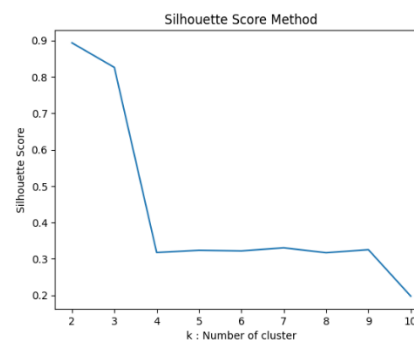
Cluster	Nilai	Selisih
1	334954.27751196176	
2	122929.31463414631	212024.9629
3	72948.50746268658	49980.80717
4	57366.96414131372	15581.54332
5	45212.33147094677	12154.63267
6	39490.71652465002	5721.614946
7	34326.33147094677	5164.385054
8	28744.514561115666	5581.81691
9	24213.67869980487	4530.835861
10	21753.59069653653	2460.088003

Dari Tabel 3 dan Tabel 4 diperoleh *cluster* yang terbaik merupakan *cluster* 2 karena setelah *cluster* 2 tidak terjadi lagi penurunan maupun kenaikan nilai

yang signifikan atau kita dapat melihat nilai selisih terbesar antara *cluster* dan didapatkan hasil yang sama yaitu: *cluster* 2 pada semester ganjil adalah *cluster* yang terbaik dengan nilai selisih terbesar sebanyak 79963.30076, diikuti dengan *cluster* 2 pada semester genap dengan nilai sebanyak 212024.9629. Setelah itu masih terdapat satu metode yang masih belum dilakukan yaitu metode *Silhouette Coefficient*, hasil grafik nilai metode *Silhouette Coefficient* dapat ditunjukkan pada Gambar 21 dan Gambar 22.



Gambar 21. *Silhouette Coefficient* Ganjil



Gambar 22. *Silhouette Coefficient* Genap

Pada Gambar 21 dan Gambar 22 tampak terlihat pada grafik, *cluster* 2 memiliki nilai yang tertinggi sehingga dapat dikatakan *cluster* yang terbaik merupakan *cluster* 2. Untuk nilai lebih jelasnya dapat dilihat pada Tabel 5 dan Tabel 6, dengan perolehan nilai *Silhouette Coefficient* sebesar 0.8171218916694154 untuk semester ganjil dan 0.8932255100239208 nilai *Silhouette Coefficient* untuk semester genap.

Tabel 5. Nilai *Silhouette Coefficient* Ganjil

Cluster	Nilai
2	0.8171218916694154
3	0.768752136606277
4	0.3163028233692416
5	0.3181276591752092
6	0.315723555531755
7	0.3184905756736919
8	0.3156090184043263
9	0.22059286959707694
10	0.21321143157238864

Tabel 6. Nilai *Silhouette Coefficient* Genap

Cluster	Nilai
2	0.8932255100239208
3	0.8259154279570559
4	0.3176174996494511
5	0.3236665007997584
6	0.3219690877410942
7	0.33032932756217553
8	0.317016103698448
9	0.3252908275073073
10	0.19760059398326962

Setelah melakukan evaluasi terhadap pemodelan dan ternyata pemodelan dirasa cukup baik, maka akan dilanjutkan proses penyebaran yang dimana dapat diserahkan berupa hasil atau laporan kepada pihak terkait. Pada penelitian ini penyebaran dilakukan khususnya kepada pihak SMKN 1 Jakarta beserta pengajar yang terkait berupa hasil atau laporan terkait hasil analisis efektivitas nilai siswa sesudah pandemi COVID-19 untuk dijadikan informasi yang dapat meningkatkan nilai siswa dalam menentukan metode pembelajaran yang tepat.

5. KESIMPULAN DAN SARAN

Dari rangkaian penelitian yang telah dilakukan, dapat diambil beberapa kesimpulan. Pertama, penerapan metode *clustering* pada data nilai siswa SMKN 1 Jakarta berhasil menghasilkan dua *cluster* pada kedua *dataset*, yakni *cluster* 0 dan *cluster* 1. *Cluster* 0 menggambarkan kelompok dengan pencapaian nilai yang baik, terdiri dari 200 siswa pada semester ganjil dan 205 siswa pada semester genap. Sementara *cluster* 1 mengandung siswa-siswa dengan pencapaian nilai kurang baik, terdiri dari 9 siswa pada semester ganjil dan 4 siswa pada semester genap. Kedua, evaluasi model *clustering* menggunakan metode *Elbow* dan *Silhouette Coefficient* menunjukkan bahwa *cluster* 2 merupakan *cluster* yang optimal, dengan perbedaan terbesar antara *cluster* pada semester ganjil sebesar 79963.30076 dan pada semester genap sebesar 212024.9629 menurut *Elbow Method*. Sementara *Silhouette Coefficient* juga mengonfirmasi *cluster* 2 sebagai *cluster* terbaik dengan nilai tertinggi, yaitu 0.8171218916694154 pada semester ganjil dan 0.8932255100239208 pada semester genap. Terakhir, hasil penelitian menyarankan peningkatan efektivitas nilai siswa setelah masa pandemi COVID-19 sebanyak 2.5%, setara dengan penambahan 5 siswa dalam *cluster* 0 yang mencerminkan pencapaian nilai baik. Hasil penelitian ini mengindikasikan bahwa masih terdapat potensi peningkatan dan penyempurnaan dalam kerangka penelitian selanjutnya. Penelitian ini menggambarkan potensi perbaikan lebih lanjut dan merekomendasikan beberapa saran untuk penelitian

mendatang. Saran-saran tersebut termasuk eksplorasi metode *clustering* alternatif dalam konteks yang sama, eksperimen dengan metode evaluasi yang berbeda, dan penggunaan *dataset* yang lebih luas guna mendapatkan analisis *cluster* yang lebih mendalam dan akurat.

DAFTAR PUSTAKA

- [1] "Kementerian Pendidikan dan Kebudayaan," 17 March 2020. [Online]. Available: <https://www.kemdikbud.go.id/main/blog/2020/03/se-mendikbud-pembelajaran-secara-daring-dan-bekerja-dari-rumah-untuk-mencegah-penybaran-covid19>
- [2] WHO, "Archived: WHO Timeline - Covid-19," 27 April 2020. [Online]. Available: <https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19>
- [3] WHO, "Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic," who, 5 May 2023. [Online]. Available: [https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-coronavirus-disease-\(covid-19\)-pandemic](https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic)
- [4] A. N. A. Ansori, "Soal Pencabutan Status Kedaruratan COVID-19 di RI, Kemenkes: Masih Dibahas," liputan6, 17 May 2023. [Online]. Available: <https://www.liputan6.com/health/read/5289370/soal-pencabutan-status-kedaruratan-covid-19-di-ri-kemenkes-masih-dibahas>
- [5] P. N. Tan, M. Steinbach and V. Kumar, Introduction to Data Mining, Pearson Addison Wesley, 2006
- [6] J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2011
- [7] I. H. Witten, E. Frank and e. a. , Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2016
- [8] KBBI, "KBBI Daring," 2016. [Online]. Available: <https://kbbi.kemdikbud.go.id/entri/efektivitas>
- [9] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining," *Journal of Data Warehousing*, 5, pp. 13-22, 2000
- [10] S. M. Tagaram, "An Overview on Clustering Methods," *IOSR Journal of Engineering*, pp. 719-725, 2012
- [11] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Journal of Computational and Applied Mathematics*, 1987.