

# PENERAPAN ALGORITMA K-MEANS CLUSTERING DALAM PENGELOMPOKAN HASIL BELAJAR SISWA DI SMP BUDI LUHUR SAMARINDA

Khoirulli Nurul Fatimah<sup>1</sup>, Heny Pratiwi<sup>2</sup>, Eka Arriyanti<sup>3</sup>

Program Studi Teknik Informatika, STMIK Widya Cipta Dharma  
Jl. M. Yamin No.25, Samarinda, 75123  
khoirullinurulf@gmail.com

## ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan hasil belajar siswa di SMP Budi Luhur Samarinda guna membantu siswa dalam memilih jurusan yang sesuai di jenjang SMA berdasarkan potensi dan kemampuan akademik mereka. Dengan menerapkan algoritma K-Means Clustering dalam metode Knowledge Discovery in Databases (KDD), diharapkan siswa dapat memperoleh panduan yang lebih tepat untuk menentukan jurusan dan mengoptimalkan prestasi akademik mereka.

Metode penelitian ini melibatkan tahap-tahap Knowledge Discovery in Databases (KDD), yaitu pembersihan data, integrasi data, seleksi data, transformasi data, penambangan data (data mining), interpretasi, dan evaluasi. Data yang digunakan terdiri dari nilai akademik 76 siswa SMP Budi Luhur Samarinda, mencakup nilai rata-rata keseluruhan mata pelajaran, rata-rata mata pelajaran eksak dan non-eksak, yang diolah menggunakan Python untuk analisis dan Streamlit untuk visualisasi serta interaksi pengguna.

Hasil analisis menunjukkan bahwa jumlah cluster optimal adalah dua, dengan nilai silhouette score sebesar 0,56. Cluster pertama terdiri dari 35 siswa (46%) dengan nilai rata-rata eksak dan non-eksak yang lebih rendah, sedangkan cluster kedua terdiri dari 41 siswa (54%) dengan nilai rata-rata yang lebih tinggi. Temuan ini diharapkan dapat memberikan wawasan yang lebih jelas mengenai potensi akademik siswa dan mendukung keputusan mereka dalam memilih jurusan di SMA.

**Kata Kunci :** *Clustering, K-Means, Hasil Belajar Siswa*

---

## 1. PENDAHULUAN

Tren yang terjadi di SMP Budi Luhur menunjukkan adanya peningkatan jumlah peserta didik seiring berjalannya waktu. Hal ini dapat disebabkan oleh berbagai faktor, seperti pertumbuhan populasi di daerah sekitar sekolah, peningkatan kesadaran akan pentingnya pendidikan, atau reputasi baik sekolah dalam memberikan pendidikan berkualitas. Meskipun pertumbuhan jumlah siswa merupakan indikator positif atas keberhasilan sekolah dalam menarik minat calon siswa, namun hal ini juga membawa tantangan tersendiri. Salah satu tantangan utama yang dihadapi adalah kesulitan bagi para siswa yang hendak lulus untuk memilih jurusan yang sesuai untuk melanjutkan ke jenjang SMA.

Dengan semakin banyaknya pilihan jurusan dan program studi di tingkat SMA, siswa seringkali merasa bingung dan terbebani dalam mengambil keputusan yang tepat. Faktor ini dapat menjadi sumber stres dan kekhawatiran bagi siswa, terutama karena pilihan jurusan yang tepat akan memengaruhi jalur pendidikan dan karir mereka di masa depan. Kesulitan dalam memilih jurusan SMA juga dapat berdampak pada kualitas pendidikan secara keseluruhan. Jika siswa tidak dapat memilih jurusan yang sesuai dengan minat, bakat, dan tujuan karir mereka, mereka mungkin akan kurang termotivasi dalam

belajar. Hal ini dapat mengakibatkan penurunan kinerja akademik, peningkatan tingkat putus sekolah, dan berpotensi mengurangi kualitas lulusan yang dihasilkan oleh sekolah.

Oleh karena itu, penting bagi SMP Budi Luhur untuk mengatasi tantangan ini dengan menyediakan bimbingan dan konseling yang memadai kepada siswa dalam proses pemilihan jurusan. Selain itu, pendekatan yang holistik dan berbasis data dalam memberikan informasi tentang berbagai pilihan jurusan dapat membantu siswa membuat keputusan yang lebih terarah dan sesuai dengan kebutuhan mereka. Dengan demikian, diharapkan dapat mengurangi tingkat kebingungan dan stres siswa serta meningkatkan kualitas pendidikan dan kelulusan di SMP Budi Luhur Samarinda.

Dari permasalahan diatas, implementasi data mining dapat diterapkan untuk dijadikan Solusi dalam mengatasi masalah tersebut. Implementasi dilakukan karena dapat menjelaskan cara kerja data mining pada objek penelitian yang akan dilakukan. Data mining digunakan sebagai indikator utama pada penelitian ini karena proses mengidentifikasi masalah yang valid dan dinilai lebih efektif dalam menyelesaikan masalah, terutama dalam mencari pola baru yang sebelumnya tidak terdeteksi.

Ada berbagai jenis metode data mining. Pada proposal ini, metode yang digunakan menggunakan clustering. Penggunaan metode clustering agar dapat dimanfaatkan sebagai segmentasi data yang berguna untuk memprediksi dan menganalisa permasalahan, terutama pada pengelompokan hasil belajar siswa di objek proposal. Jenis clustering yang digunakan adalah K-Means. K-Means sangat umum digunakan sebagai teknik clustering. Metode clustering K-Means diterapkan pada proposal ini karena dinilai cukup mudah untuk dipahami dan diterapkan. Selain itu, proses pembelajaran pada metode ini relatif lebih cepat.

Penelitian yang dilakukan berkaitan dengan pengelompokan nilai. Pengelompokan disini merujuk pada proses mengelompokkan objek atau data ke dalam kelompok-kelompok yang memiliki kesamaan tertentu, tanpa adanya label kategori sebelumnya yang akan diterapkan menggunakan metode clustering K-Means. Data-data yang digunakan pada objek penelitian ini berupa nilai akademik (nilai rapor) selama 5 semester karena sebagai acuan ke jenjang selanjutnya.

Sistem yang digunakan dalam penelitian ini berbasis web. Sistem ini diharapkan mampu mempermudah user dalam melihat hasil belajar siswa SMP Budi Luhur Samarinda. Penerapan algoritma k-means clustering diharapkan dapat membantu siswa SMP Budi Luhur Samarinda dalam pemilihan jurusan ke jenjang selanjutnya.

## 2. RUANG LINGKUP PENELITIAN

### 2.1 RUMUSAN MASALAH

Berdasarkan latar belakang di atas, didapatkan rumusan masalah dalam penelitian ini yaitu, “Bagaimana penerapan algoritma K-Means Clustering dalam pengelompokan hasil belajar siswa di SMP Budi Luhur Samarinda?”.

### 2.2 BATASAN MASALAH

Pada penelitian ini, dapat diidentifikasi beberapa batasan dari masalah yang diantaranya yaitu sebagai berikut:

1. Penelitian ini menggunakan data kelas IX tahun ajaran 2023/2024 yang tersedia di SMP Budi Luhur Samarinda.
2. Hasil belajar siswa akan diukur melalui data akademik pada rapor yang tersedia, bukan melalui survei atau wawancara langsung dengan siswa.
3. Variabel clustering terdiri dari rerata keseluruhan nilai akademik, rerata nilai mata pelajaran eksak, dan rerata nilai mata pelajaran non eksak.

### 2.3 TUJUAN PENELITIAN

Penelitian ini bertujuan untuk memberikan rekomendasi jurusan kepada siswa SMP Budi Luhur Samarinda kelas IX yang hendak lulus dan melanjutkan ke jenjang selanjutnya dengan mengelompokkan siswa berdasarkan hasil belajarnya, serta membantu sekolah untuk meningkatkan kualitas lulusan.

## 3. BAHAN DAN METODE

Adapun bahan dan metode yang gunakan dalam membangun penelitian ini yaitu:

### 3.1 Penerapan

Menurut Kamus Besar Bahasa Indonesia (KBBI), pengertian penerapan adalah perbuatan menerapkan,

sedangkan menurut beberapa ahli, penerapan adalah suatu perbuatan mempraktekkan suatu teori, metode, dan hal lain untuk mencapai tujuan tertentu dan untuk suatu kepentingan yang diinginkan oleh suatu kelompok atau individu bahkan golongan yang telah terencana dan tersusun sebelumnya.

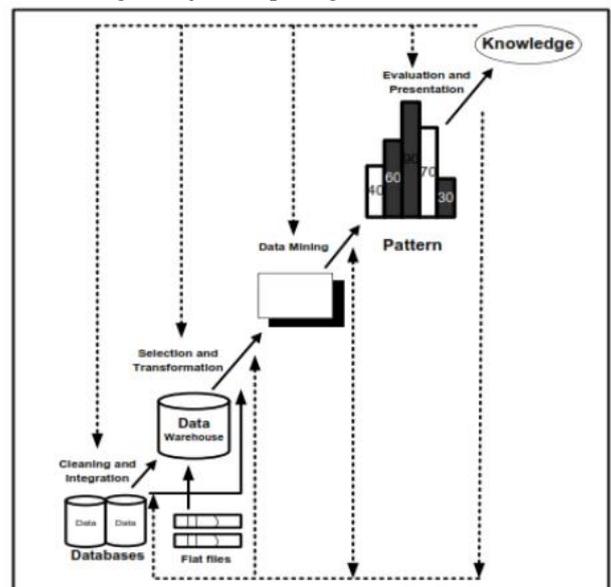
Menurut (Putri, 2019) Penerapan adalah proses, cara atau perbuatan sebagai kemampuan meningkatkan bahanbahan yang dipelajari dengan rencana yang telah disusun secara sistematis, seperti metode, konsep dan teori.

Berdasarkan pengertian-pengertian di atas, dapat disimpulkan bahwa kata penerapan (implementasi) bermuara pada aktifitas, adanya aksi, tindakan, atau mekanisme suatu sistem. Ungkapan mekanisme mengandung makna bahwa penerapan (implementasi) bukan sekedar aktifitas, tetapi cara yang dilakukan dalam kegiatan agar dapat mencapai tujuan yang diinginkan. Adapun unsur-unsur penerapan meliputi:

1. Adanya program yang akan dilaksanakan.
2. Adanya kelompok target, yaitu masyarakat yang menjadi sasaran dan
3. diharapkan akan menerima manfaat dari program tersebut.
4. Adanya pelaksanaan, baik organisasi atau perorangan yang bertanggung jawab dalam pengelola, pelaksanaan maupun pengawasan dari proses penerapan tersebut.

### 3.2 Data Mining

Data mining adalah sebuah untai proses, maka terbagi menjadi beberapa tahap. Tahapan tersebut akan bersifat interaktif, pengguna akan terlibat langsung atau dengan perantara KDD (Untari, 2014). Berikut tahapan data mining ditunjukkan pada gambar 2.1



Gambar 2.1. Tahapan Data Mining

Sumber : (xerma, 2015)

Tahapan data mining dibagi menjadi bagian-bagian yaitu :

1. Pembersihan data (data cleaning)  
Data cleaning (pembersihan data) adalah proses yang bertujuan untuk menghilangkan data yang

mengandung error, nilai suatu atribut yang tidak benar yang menyimpang dari yang diharapkan (noise) dan data yang tidak konsisten untuk dibuang dari koleksi data dengan tujuan agar tidak mengganggu proses data mining nantinya.

2. Penggabungan data (data integration)  
Data integration (penggabungan data) adalah proses yang bertujuan untuk menggabungkan data-data yang berasal dari berbagai sumber data yang dikumpulkan.
3. Seleksi data (data selection)  
Data selection (seleksi data) adalah proses yang berfungsi untuk mengambil data dari database yang relevan dengan tugas analisis.
4. Transformasi data (data transformation)  
Data transformation (transformasi data) berfungsi mentransformasikan dan mengkonsolidasikan data ke dalam bentuk yang sesuai untuk penambangan dengan melakukan operasi ringkasan atau agregasi.
5. Data mining  
Data mining yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Pemilihan metode yang tepat sangat bergantung pada tujuan dan proses data mining.
6. Evaluasi pola (pattern evaluation)  
Pattern evaluation (evaluasi pola) berfungsi untuk mengidentifikasi pola-pola yang mewakili pengetahuan berdasarkan ukuran-ukuran yang menarik.
7. Presentasi pengetahuan (knowledge presentation)  
Knowledge presentation (presentasi pengetahuan) digunakan untuk memvisualisasikan dan merepresentasikan pengetahuan yang telah ditambang.

Berdasarkan pengertian di atas tentang data mining dapat disimpulkan bahwa, data mining merupakan proses untuk menemukan suatu pola menarik dan pengetahuan dari suatu basis data yang besar. Pada penelitian ini data mining dilakukan untuk melakukan proses pengelompokan hasil belajar siswa berdasarkan nilai akademik.

### 3.3 Algoritma K-Means

K-Means merupakan Algoritma clustering yang pertama kali diperkenalkan oleh James B MacQueen pada tahun 1976. Metode ini merupakan suatu metode clustering non-heirarchial yang umum digunakan yang relatif sederhana untuk mengelompokkan data dalam jumlah besar.

K-Means merupakan metode klasterisasi yang sering digunakan diberbagai bidang karena penggunaannya sederhana, mudah untuk di implementasikan, mampu untuk mengklaster data yang besar. Algoritma K-Means merupakan metode berbasis jarak yang membagi data kedalam sejumlah cluster dan dalam setiap tahapan tertentu setiap objek harus masuk dalam kelompok, pada tahap selanjutnya objek dapat berpindah ke kelompok lain. Dalam algoritma ini pada dasarnya melakukan proses clustering tetapi tergantung dari data yang didapat dan konklusi yang dicapai.

Maka dari itu algoritma K-Means mempunyai aturan yaitu ada jumlah cluster yang akan diinputkan

dan hanya dapat memiliki atribut yang bertipe numerik. Pada awalnya dalam algoritma K-Means melakukan pengambilan sebagian dari banyaknya populasi untuk dijadikan cluster awal.

Ada banyak cara dalam memberi nilai awal misalnya dengan pengambilan data sampel awal dari objek. Pusat cluster dipilih secara acak yang berada dari beberapa populasi data. Setelah mendapatkan pusat cluster awal, algoritma K-Means melakukan pengujian masing-masing komponen ke salah satu pusat cluster yang telah didefinisikan jarak minimumnya antar komponen dengan tiap-tiap pusat cluster.

Posisi pusat cluster akan melakukan perhitungan kembali sampai semua komponen data dapat digolongkan kedalam setiap cluster dan akan membentuk posisi cluster baru.

Dalam penerapannya, k-means memisahkan data dengan melakukan perulangan secara terus-menerus sampai tidak ada perubahan data dalam setiap segmentasi. Langkah-langkah dalam metode k-means adalah sebagai berikut:

1. Tentukan nilai k sebagai jumlah klaster yang ingin dibentuk.
2. Bangkitkan k centroid (titik pusat cluster) awal.
3. Hitung jarak setiap data ke masing-masing centroid menggunakan rumus Manhattan Distance seperti pada persamaan berikut.

$$D(X,Y) = \sum |X_i - Y_i|$$

Keterangan :

$D(X,Y)$  = jarak objek antara  $X_i$  dan  $Y_i$

$X_i$  = Koordinat dari objek  $X_i$  pada dimensi  $i$

$Y_i$  = Koordinat dari objek  $Y_i$  pada dimensi  $i$

4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya.
5. Tentukan centroid baru dengan cara menghitung nilai rata-rata dari data-data yang ada pada centroid yang sama dengan rumus seperti persamaan berikut.  
 $C_i = ((\sum x)/n)$
6. Dimana n adalah banyaknya dokumen dalam cluster  $i$  dan x adalah data yang akan dihitung.
7. Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama.

### 3.4 Clustering

Menurut Yulianto (2022), clustering merupakan suatu metode pengelompokan data. Lebih jelasnya, clustering adalah metode pengelompokan data yang digunakan untuk mengenali kelompok-kelompok (cluster) yang dihasilkan dari pengelompokan unsur-unsur yang lebih kecil berdasarkan adanya kemiripan satu sama lain. Kemiripan yang menjadi dasar pengelompokan tidak bersifat universal sehingga ukuran-ukuran penyamannya harus dijabarkan terlebih dahulu oleh peneliti atau penganalisis.

Clustering merupakan metode pengelompokan data yang sering digunakan sebagai satu di antara metode data mining atau penggalian data. Clustering adalah proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster. Maka itu, metode clustering ini berguna untuk menemukan kelompok yang tidak dikenal dalam data.

Dalam business intelligence, clustering bisa mengelompokkan banyak customer ke beberapa kelompok, contohnya mengelompokkan customer ke dalam beberapa cluster dengan kesamaan karakteristik yang kuat.

*Clustering* atau klusterisasi adalah salah satu alat bantu pada data mining yang bertujuan mengelompokkan objek-objek ke dalam cluster-cluster. Cluster adalah sekelompok atau sekumpulan objek-objek data yang similar satu sama lain dalam cluster yang sama dan dissimilar terhadap objek-objek yang berbeda cluster. Objek akan dikelompokkan ke dalam satu atau lebih cluster sehingga objek-objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan lainnya. Clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan data yang lain (Gifthera Dwilestari, Mulyawan, Martanto, Irfan Ali, 2021).

*Clustering* merupakan salah satu metode data mining yang bersifat tanpa arahan (unsupervised), maksudnya metode ini diterapkan tanpa adanya latihan (training) dan tanpa ada guru serta tidak memerlukan target output. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu hierarchical clustering dan non-hierarchical clustering (Yuni Radana Sembiring, Saifullah, Riki Winanjaya, 2021).

### 3.5 Tujuan K-Means Clustering

Menurut Salsabila (2019), tujuan pengelompokan (clustering) data dapat dibedakan menjadi 2 :

1. Pengelompokan untuk pemahaman : Kelompok yang terbentuk harus menangkap struktur alami data, biasanya proses pengelompokan dalam tujuan ini hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan inti seperti peringkasan atau summarization (rata-rata), pelabelan kelas pada setiap kelompok untuk kemudian digunakan sebagai data latih klasifikasi dan sebagainya.
2. Pengelompokan untuk penggunaan : mencari prototype kelompok yang paling representatif terhadap data, memberikan abstraksi dari setiap objek data dalam kelompok dimana sebuah data terletak di dalamnya.

### 3.6 Kelebihan K-Means Clustering

Menurut Salsabila (2019), kelebihan dari algoritma K-Means:

1. Mudah diterapkan dan dijalankan
2. Mudah diadaptasi
3. Memiliki tingkat konvergensi yang tinggi
4. Menghasilkan cluster yang lebih padat jika dibandingkan dengan metode hirarki.

### 3.7 Kelemahan K-Means Clustering

Menurut Hakim (2020), metode ini memiliki beberapa kelemahan yaitu :

1. Jumlah cluster sebanyak K harus ditentukan sebelum dilakukan perhitungan.
2. Bila jumlah data tidak terlalu banyak, maka mudah dalam menentukan centroid awal.

3. Kontribusi dari atribut dalam pengelompokan tidak diketahui karena semua atribut dianggap mempunyai bobot yang sama.
4. Tidak pernah mengetahui real cluster dengan menggunakan data yang sama, tetapi dimasukkan dengan cara yang berbeda, mungkin dapat memproduksi cluster yang berbeda jika jumlah datanya sedikit.

### 3.8 Pengelompokan

Menurut (Tan, Steinbach, & Kumar, 2014), Pengelompokan atau clustering adalah proses mengatur data ke dalam kelompok yang masing-masing kelompok terdiri dari objek yang lebih mirip satu sama lain dibandingkan dengan objek di kelompok lainnya Tujuan pekerjaan pengelompokan (clustering) data dapat dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan pengelompokan untuk penggunaan.

Jika tujuannya untuk pemahaman, kelompok yang terbentuk harus menangkap struktur alami data, biasanya proses pengelompokan dalam tujuan ini hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan inti seperti peringkasan atau summarization (rata-rata, standar deviasi), pelabelan kelas pada setiap kelompok untuk kemudian digunakan sebagai data latih klasifikasi dan sebagainya. Sementara jika untuk penggunaan, tujuan utama pengelompokan biasanya adalah mencari prototipe kelompok yang paling representative terhadap data, memberikan abstraksi dari setiap objek data dalam kelompok di mana sebuah data terletak di dalamnya

### 3.9 Hasil Belajar

Hasil belajar menurut Amir & Risnawati (2015) adalah kemampuan yang diperoleh anak setelah melalui kegiatan belajar. Menurut Suprijono (2015) hasil belajar adalah "pola-pola perilaku, nilai-nilai, pengertian-pengertian, sikap-sikap, apresiasi dan keterampilan". Merujuk pemikiran Gagne, hasil belajar berupa :

1. Informasi verbal yaitu kapabilitas mengungkapkan pengetahuan dalam bentuk Bahasa, baik lisan maupun tertulis. Kemampuan merespon secara spesifik terhadap rangsangan spesifik. Kemampuan tersebut tidak memerlukan manipulasi symbol, pemecahan masalah maupun penerapan aturan.
2. Keterampilan intelektual yaitu kemampuan mempresentasikan konsep dan lambing. Keterampilan intelektual terdiri dari kemampuan mengategorisasi, kemampuan analisis-sintesis fakta-konsep dan mengembangkan prinsip-prinsip keilmuan. Keterampilan intelektual merupakan kemampuan melakukan aktivitas kognitif bersifat khas.
3. Strategi kognitif yaitu kecakapan menyalurkan dan mengarahkan aktivitas kognitifnya sendiri. Kemampuan ini meliputi penggunaan konsep kaidah dalam pemecahan masalah.
4. Keterampilan motoric yaitu kemampuan melakukan serangkaian gerak jasmani dalam urusan kordinasi, sehingga terwujud otomatisme gerak jasmani.
5. Sikap adalah kemampuan menerima atau menolak objek berdasarkan penilaian terhadap objek tersebut. Sikap berupa kemampuan menginternalisasikan dan eksternalisasikan nilai-nilai. Sikap merupakan

kemampuan menjadikan nilai-nilai sebagai standar perilaku.

Berdasarkan pendapat di atas dapat disimpulkan bahwa hasil belajar dapat didefinisikan sebagai kemampuan yang dimiliki siswa setelah melakukan aktivitas pembelajaran dalam bentuk angka atau skor yang diperoleh dari penilaian atau tes yang dilaksanakan dalam proses yang sedang berlangsung.

### 3.10 Python

Python adalah bahasa pemrograman tingkat tinggi yang populer untuk analisis data dan pemodelan. Dikembangkan pada tahun 1990 oleh Guido van Rossum, Python telah menjadi alat yang populer di berbagai bidang, termasuk ilmu data, pemrosesan bahasa alami, dan pemodelan epidemiologi. Python memiliki sintaks yang sederhana dan mudah dipahami, sehingga membuatnya cocok untuk pemula dan pengguna yang berpengalaman. Python memiliki berbagai library yang berguna untuk analisis data antara lain pandas, numpy, matplotlib, dan seaborn (Hermanto, Salim, Wu, Salim, & Gunadi, 2023).

### 3.11 Streamlit

Streamlit adalah sebuah framework open-source yang digunakan untuk membangun antarmuka pengguna (UI) interaktif untuk aplikasi data science. Framework ini dirancang khusus untuk mempermudah pengembangan aplikasi web dengan menggunakan Python. Dalam konteks seorang data scientist, streamlit memungkinkan mereka untuk dengan cepat membuat dan berbagi aplikasi interaktif untuk memvisualisasikan data, menjalankan model machine learning, dan mengeksplorasi hasil analisis data, dengan Streamlit, peneliti dapat dengan mudah membuat antarmuka pengguna yang responsif dan dinamis tanpa perlu menulis banyak kode (DQLab, 2023).

### 3.12 Visual Code Studio

Visual Code Studio merupakan code editor gratis yang dapat dijalankan pada perangkat desktop dengan basis Linux, Windows, dan Mac OS yang mana dikembangkan oleh Microsoft. Aplikasi visual studio code adalah sebuah software editor powerful namun tetap ringan saat digunakan. Aplikasi ini juga dapat digunakan untuk membuat serta mengedit source code memakai berbagai bahasa pemrograman seperti Node.js, TypeScript, dan JavaScript. Bahkan untuk visualnya tergolong kompatibel dengan bahasa serta runtime environment lainnya seperti .NET, Java, Python, dan PHP (Studio, 2023).

### 3.13 Metode Pengujian Sistem

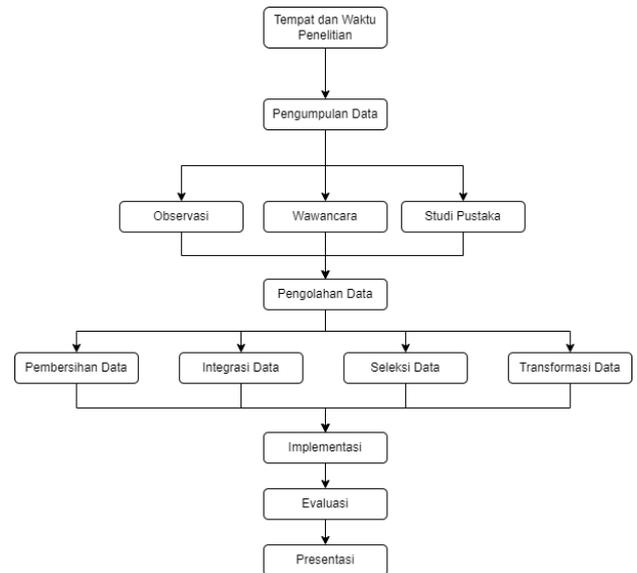
Pada data mining dilakukan pengujian validasi dengan cara membandingkan sistem yang telah dibangun dengan dataset yang sudah diuji menggunakan metode K-Means clustering. Selain itu, dilakukan suatu pengujian fungsional sistem dengan blackbox testing.

Menurut Hakim (2020), Blackbox testing merupakan pengujian yang berpusat pada kebutuhan fungsional perangkat lunak yang memungkinkan untuk memperoleh sekumpulan kondisi input yang secara penuh memeriksa fungsional dari sebuah aplikasi. Blackbox testing berusaha menemukan kesalahan-kesalahan seperti kesalahan fungsi dan kesalahan tampilan aplikasi. Pengujian sistem dengan blackbox ini juga dilakukan untuk mengetahui apabila terjadi kesalahan atau

kekurangan sehingga sistem yang dibuat dapat diperbaiki untuk menghasilkan sistem sesuai dengan yang diharapkan. Selain blackbox testing, pengujian sistem juga menggunakan *software rapidminer* untuk membandingkan hasil yang diperoleh.

## 4. RANCANGAN SISTEM ATAU APLIKASI

Pada tahap rancangan sistem / aplikasi ini bertujuan untuk memberikan suatu gambaran umum tentang tahapan *Knowledge Discovery in Database (KDD)*.



Gambar 4.1 Tahapan Penelitian

Informasi yang dikumpulkan perlu melalui suatu proses tertentu untuk menghasilkan kejelasan atau kesimpulan yang sesuai dengan tujuan pengumpulan data atau alur penelitian yang akan dilakukan. Dalam konteks penelitian ini, metode yang diterapkan adalah metode *Knowledge Discovery in Database (KDD)*. Data mining memiliki tujuan utama dalam menemukan, menggali atau menambang pengetahuan dari data atau informasi yang dimiliki. Metode ini mempunyai tahapan-tahapan yang jelas dan setiap tahapannya harus diselesaikan terlebih dahulu untuk menghindari terjadinya pengulangan dalam tahapan sehingga pengembangan sistem yang dilakukan dapat memperoleh hasil yang diinginkan, berikut tahapan-tahapan tersebut :

### 4.1 Pemahaman Data

Berdasarkan penelitian yang dilakukan didapatkan pemahaman terhadap data yang didapatkan berupa nilai siswa. Pada data yang didapatkan terdapat dua puluh dua atribut pada data nilai siswa, diantaranya ada nomor peserta ujian, nama, tempat tanggal lahir, nama orang tua, NIS, NISN, nilai matematika, ilmu pengetahuan alam, Pendidikan agama islam, Pendidikan Pancasila dan kewarganegaraan, Bahasa Indonesia, ilmu pengetahuan sosial, Bahasa Inggris, seni budaya, Pendidikan jasmani, olahraga dan Kesehatan, prakarya, muatan lokal, jumlah rata-rata semester 1-6, rata-rata keseluruhan, rata-rata mata Pelajaran eksak, dan rata-rata mata Pelajaran non eksak. Kurikulum yang digunakan untuk siswa kelas IX pada tahun ajaran 2023/2024 adalah Kurikulum K13, di

mana nilai berupa angka, dan data yang diproses ini juga berupa angka.

## 4.2 Pengolahan Data

### 4.2.1 Pembersihan Data

Pada tahap ini dilakukan pembersihan data dari kesalahan, duplikasi, nilai yang hilang, dan memastikan semua data dalam format yang konsisten yaitu numerik di seluruh kolom nilai pelajaran.

#### 1. Membaca Dataset seperti gambar 4.2

```
file_path = 'nilai_siswa.xlsx'
df = pd.read_excel(file_path)
df
```

Gambar 4.1 Membaca Dataset

Dataset ini berisi 76 baris dan 22 kolom.

#### 2. Mengecek tipe data pada kolom dataset seperti gambar 4.3

```
df.info()
class pandas.core.frame.DataFrame
Kategori: 76 entries x 19 to 75
Data columns (total 22 columns):
 #   Column                Non-Null Count  Dtype
---  ---                ---
 0   NO                     76 non-null    int64
 1   NAMA PESERTA UJIAN SEKOLAH 76 non-null    object
 2   NAMA                   76 non-null    object
 3   TEMPAT, TGL LAHIR     76 non-null    object
 4   NAMA ORANG TUA        76 non-null    object
 5   NIS                    76 non-null    int64
 6   NISN                   76 non-null    int64
 7   Matematika            76 non-null    float64
 8   Ilmu Pengetahuan Alam 76 non-null    float64
 9   Pendidikan Agama dan Budi Pekerti 76 non-null    float64
10   Pendidikan Pancasila dan Kewargabangsaan 76 non-null    float64
11   Bahasa Indonesia      76 non-null    float64
12   Ilmu Pengetahuan Sosial 76 non-null    float64
13   Bahasa Inggris        76 non-null    float64
14   Seni Budaya           76 non-null    float64
15   Pendidikan Jaman, Olahraga dan Kesehatan 76 non-null    float64
16   Prakarya dan atau Informatika 76 non-null    float64
17   Bahasa Lokal          76 non-null    float64
18   RATA RATA MAPEL EKSAK 76 non-null    float64
19   RATA RATA KESELURUHAN 76 non-null    float64
20   RATA RATA MAPEL NON EKSAK 76 non-null    float64
21   RATA RATA MAPEL NON EKSAK 76 non-null    float64
dtypes: float64(17), int64(3), object(2)
```

Gambar 4.3 Mengecek Tipe Data

Terdapat 15 kolom bertipe *float64*, 3 kolom bertipe *integer*, dan 4 kolom bertipe *object*.

#### 3. Mengecek missing values seperti gambar 4.4.

```
# Mengecek baris yang diduplikasi
duplicate_rows = df[df.duplicated()]
print("Terdapat {duplicate_rows.shape[0]} baris yang diduplikasi")
print(duplicate_rows)

# Mengecek nilai yang hilang di setiap kolom
missing_values = df.isnull().sum()
print("Terdapat nilai yang hilang di setiap kolom:")
print(missing_values)

df
```

Gambar 4.4 Mengecek Missing Values

Dari hasil diatas tidak terdapat data yang duplikat, maupun data yang kosong.

### 4.2.2 Seleksi Data

Pada tahap ini akan dilakukan pemilihan kolom-kolom yang relevan untuk analisis clustering, yaitu kolom rata-rata keseluruhan, nilai eksak, dan nilai non eksak.

#### 1. Mengubah nama kolom yang dipilih yaitu kolom RATA RATA KESELURUHAN menjadi Rata-Rata, kolom RATA RATA MAPEL EKSAK menjadi Eksak, dan kolom RATA RATA MAPEL NON EKSAK menjadi Non Eksak menggunakan syntax gambar 4.5.

```
df.rename(index=str, columns={
    'RATA RATA KESELURUHAN': 'Rata-Rata',
    'RATA RATA MAPEL EKSAK': 'Eksak',
    'RATA RATA MAPEL NON EKSAK': 'Non Eksak'
}, inplace=True)
```

Gambar 4.5 Mengubah Nama Kolom

#### 2. Memilih kolom yang relevan, yaitu kolom rata-rata, eksak, dan non eksak menggunakan syntax gambar 4.6.

```
x = df[['Rata-Rata', 'Eksak', 'Non Eksak']]
x
```

Gambar 4.6 Memilih Kolom

### 4.2.3 Transformasi Data

Pada tahap ini akan memastikan bahwa semua nilai numerik berada dalam skala yang sama dengan normalisasi data.

#### 1. Transformasi data menggunakan StandardScaler() yang digunakan untuk menormalisasi fitur dengan menghapus mean dan menskalakan ke unit variance seperti pada gambar 4.7.

```
scaler = StandardScaler()
X = scaler.fit_transform(x)
```

Gambar 4.7 Transformasi Data

#### 2. Mengecek hasil data yang telah di transformasikan seperti gambar 4.8.

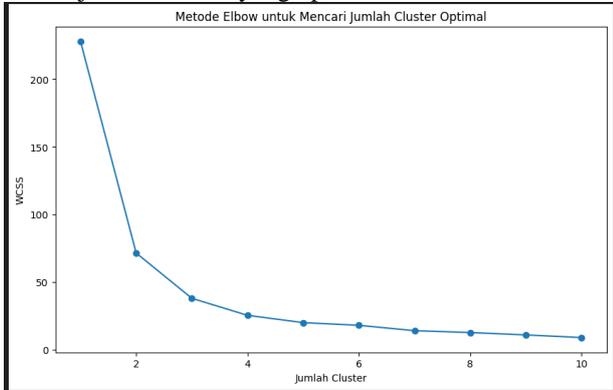
```
print(X)
```

Gambar 4.8 Mengecek Hasil Data

#### 4.2.4 Implementasi

Pada tahap ini akan menentukan jumlah *cluster* (*k*) dan menerapkan algoritma *K-means*.

1. Menentukan jumlah cluster menggunakan metode elbow seperti gambar 4.9 untuk menentukan jumlah cluster yang optimal.



Gambar 4.9 Grafik *Elbow*

Dari grafik dapat dilihat bahwa jumlah cluster optimal biasanya ditunjukkan oleh "siku" atau titik di mana penurunan Within-Cluster Sum of Squares (WCSS) mulai melambat secara signifikan. Pada grafik terlihat penurunan WCSS yang tajam terjadi antara cluster 2 dan 3, dan setelah itu penurunan WCSS mulai melambat. Ini menunjukkan bahwa setelah 3 cluster, penambahan cluster baru tidak lagi memberikan pengurangan WCSS yang signifikan, jadi berdasarkan grafik elbow ini, jumlah cluster optimal yang terbaik adalah 2 atau 3.

2. Menerapkan algoritma *k-means* clustering pada data yang sudah diproses menggunakan syntax seperti gambar 4.10.

```
k = 3
kmeans = KMeans(n_clusters=k, random_state=42)
kmeans.fit(X)
clusters = kmeans.labels_
✓ Oh.
```

Gambar 4.10 Penerapan Algoritma *K-Means*

#### 4.2.5 Evaluasi

Pada tahap ini, evaluasi hasil clustering akan dilakukan menggunakan Silhouette Score seperti gambar 4.11 untuk memastikan kualitas clustering yang digunakan. Silhouette Score berkisar antara -1 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa objek jauh lebih mirip dengan objek lain dalam cluster yang sama dibandingkan dengan objek di cluster lain, ini menunjukkan clustering yang baik. Nilai yang mendekati 0 menunjukkan bahwa objek berada di batas antara dua cluster. Nilai negatif menunjukkan bahwa objek lebih mirip dengan objek di cluster lain dibandingkan dengan cluster mereka sendiri, ini menunjukkan clustering yang buruk.

```
# Mengevaluasi berbagai jumlah cluster menggunakan Silhouette Score
for n_clusters in range(2, 11):
    kmeans = KMeans(n_clusters=n_clusters, random_state=42)
    cluster_labels = kmeans.fit_predict(X)
    silhouette_avg = silhouette_score(X, cluster_labels)
    print(f"Dari n_clusters = {n_clusters}, Rata rata nilai silhouette adalah : {silhouette_avg}")
✓ Oh.
```

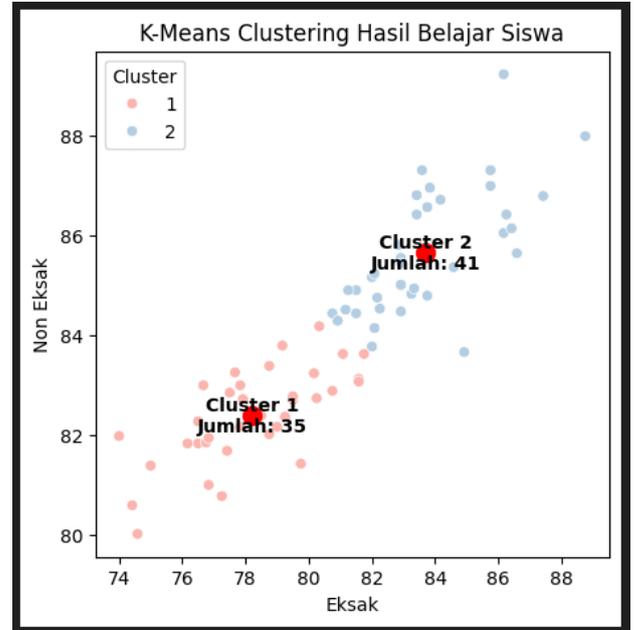
```
Dari n_clusters = 2, Rata rata nilai silhouette adalah : 0.5615379617082531
Dari n_clusters = 3, Rata rata nilai silhouette adalah : 0.5253735049683584
Dari n_clusters = 4, Rata rata nilai silhouette adalah : 0.474398031479321
Dari n_clusters = 5, Rata rata nilai silhouette adalah : 0.4235315678893906
Dari n_clusters = 6, Rata rata nilai silhouette adalah : 0.372773726953268
Dari n_clusters = 7, Rata rata nilai silhouette adalah : 0.38603878484746284
Dari n_clusters = 8, Rata rata nilai silhouette adalah : 0.34202210461438337
Dari n_clusters = 9, Rata rata nilai silhouette adalah : 0.34314948674432937
Dari n_clusters = 10, Rata rata nilai silhouette adalah : 0.380286299155283
```

Gambar 4.11 Evaluasi Menggunakan *Silhouette Score*

dari hasil evaluasi pada gambar 4.10 menunjukkan bahwa cluster yang terbaik adalah 2 dengan nilai 0,56 dibandingkan 3 dengan nilai 0,52.

#### 4.2.6 Presentasi

Pada tahap ini akan dilakukan penyajian hasil analisis dalam bentuk visualisasi menggunakan grafik plot pada gambar 4.12 yang dapat membantu dalam memahami distribusi data dan pembagian cluster hasil penerapan *k-means*.



Gambar 4.12 Visualisasi Grafik Plot

Grafik diatas menunjukkan hasil clustering dari data hasil belajar siswa dalam dua dimensi yaitu nilai "Eksak" dan nilai "Non Eksak", yang mana terdapat dua cluster yang ditandai dengan warna yang berbeda. Cluster 1 ditunjukkan dengan titik berwarna merah, sedangkan Cluster 2 ditunjukkan dengan titik berwarna biru. Sumbu X merepresentasikan nilai "Eksak" siswa, sementara sumbu Y merepresentasikan nilai "Non Eksak" siswa.

Visualisasi ini menunjukkan bahwa Siswa dalam Cluster 1 (titik merah) sebanyak 35 siswa, menunjukkan nilai yang lebih rendah dalam "Eksak" dan "Non Eksak". Siswa dalam Cluster 1 mungkin lebih cocok untuk program SMA yang menawarkan dukungan tambahan atau fokus pada pengembangan keterampilan non-akademik, seperti program kejuruan atau sekolah dengan program bimbingan yang kuat.

Siswa dalam Cluster 2 (titik biru) sebanyak 41 siswa, cenderung memiliki nilai "Eksak" dan "Non Eksak" yang lebih tinggi. Siswa ini bisa direkomendasikan untuk mengambil program SMA dengan penekanan pada mata pelajaran sains dan matematika, karena mereka sudah menunjukkan kekuatan di area ini.

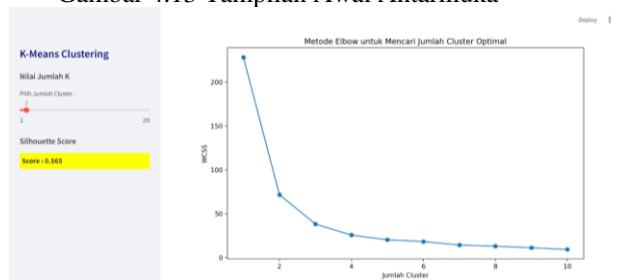
### 5. IMPLEMENTASI

Hasil penelitian ini merupakan implementasi dari perencanaan yang telah disusun pada bab sebelumnya. Berikut adalah hasil penerapan algoritma *K-Means* Clustering dalam pengelompokan hasil belajar siswa di SMP Budi Luhur Samarinda. Implementasi ini menampilkan isi dataset, slider jumlah cluster yang dipilih dan skor Silhouette yang dihasilkan seperti gambar 4.13,

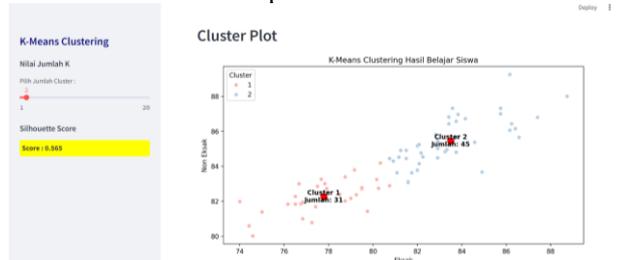
serta menyajikan metode elbow seperti gambar 4.14 yang digunakan untuk menentukan jumlah cluster optimal, dan gambar 4.15 untuk menampilkan plot cluster untuk visualisasi data yang dikelompokkan, dan tabel dataset dengan tambahan kolom cluster yang menunjukkan kelompok masing-masing data. Tujuan dari implementasi ini adalah untuk memudahkan pihak sekolah dalam menggunakan visualisasi tersebut.



Gambar 4.13 Tampilan Awal Antarmuka



Gambar 4.14 Tampilan Grafik Metode *Elbow*



Gambar 4.15 Tampilan *Cluster Plot*

## 6. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, beberapa temuan penting dalam penerapan metode K-Means Clustering terhadap hasil belajar siswa adalah sebagai berikut:

1. **Optimalitas Cluster**  
Hasil *clustering* terbaik diperoleh dengan menggunakan 2 *cluster* ( $k=2$ ), yang menunjukkan *Silhouette Score* sebesar 0,56.
2. **Pengujian dengan RapidMiner**  
Pengujian menggunakan software *RapidMiner* memberikan hasil yang konsisten dan valid dengan model *clustering* yang dirancang.
3. Setelah ditinjau kembali, data awal menunjukkan bahwa siswa dalam salah satu *cluster* memiliki nilai rata-rata eksak dan non-eksak yang lebih tinggi dibandingkan dengan siswa di *cluster* lainnya. *Cluster* 1 terdiri dari 35 siswa yang memiliki nilai rata-rata eksak lebih rendah dibandingkan dengan siswa di *Cluster* 2. *Cluster* 2 terdiri dari 41 siswa yang memiliki nilai rata-rata eksak dan non-eksak yang lebih tinggi dibandingkan dengan *Cluster* 1. Siswa dalam *cluster* 2

ini menunjukkan performa yang lebih baik dalam kedua kategori mata pelajaran, baik eksak maupun non-eksak.

4. Rekomendasi yang diberikan berdasarkan hasil *clustering* ini belum siap untuk implementasi nyata dan hanya merupakan model *clustering* awal, karena data yang didapatkan masih terbatas.

Dengan hasil ini, dapat disimpulkan bahwa *Clustering* menggunakan metode *K-Means* efektif dalam mengelompokkan siswa berdasarkan performa akademik mereka, meskipun perlu adanya pengumpulan data yang lebih komprehensif untuk implementasi yang lebih mendalam dan akurat.

## 7. SARAN

Penelitian ini masih memiliki banyak kekurangan, maka saran yang dapat disampaikan sebagai berikut.

1. **Bagi Peserta Didik**  
Peserta didik dalam *cluster* 1 mungkin lebih cocok untuk program SMA yang menawarkan dukungan tambahan atau fokus pada pengembangan keterampilan non-akademik, seperti program kejuruan atau sekolah dengan program bimbingan yang kuat, dan bagi peserta didik yang berada dalam *cluster* 2 bisa direkomendasikan untuk mengambil program SMA dengan penekanan pada mata pelajaran sains dan matematika, karena mempunyai kemampuan akademik yang tinggi.
2. **Bagi Orang Tua**  
Orang tua perlu lebih memahami potensi dan minat anak-anak mereka berdasarkan hasil *clustering* ini. Untuk anak-anak yang berada dalam *cluster* 1, orang tua bisa mempertimbangkan sekolah dengan program yang dapat membantu mengembangkan keterampilan praktis dan non-akademik. Sedangkan untuk anak-anak di *cluster* 2, orang tua dapat mendorong mereka untuk mengambil program yang menekankan mata pelajaran sains dan matematika. Selain itu, orang tua dapat berperan aktif dalam memberikan dukungan belajar di rumah dan berkomunikasi dengan pihak sekolah untuk memantau perkembangan anak.
3. **Bagi SMP Budi Luhur Samarinda**  
Melalui hasil *clustering* ini, dapat diidentifikasi siswa-siswa yang memerlukan pendekatan personal dalam konseling. Konselor sekolah dapat menggunakan informasi ini untuk memberikan rekomendasi yang lebih spesifik dan individual untuk setiap siswa, membantu mereka memilih jalur pendidikan yang sesuai dengan potensi dan minat mereka.
4. **Bagi Penelitian Selanjutnya**  
Untuk meningkatkan kinerja sistem dan melakukan penyempurnaan lebih lanjut, disarankan untuk mempertimbangkan hal-hal berikut:
  1. **Pemilihan Centroid Awal** : Pemilihan centroid awal yang acak dapat membuat hasil yang berbeda-beda. Maka disarankan untuk melakukan perhitungan secara manual atau menggunakan metode inisialisasi centroid yang lebih canggih, seperti *K-Means++* untuk memastikan hasil yang lebih akurat dan konsisten.
  2. **Pengumpulan data yang lebih beragam** dengan menambah variabel atau fitur lain yang relevan

dapat memberikan gambaran yang lebih komprehensif tentang kemampuan dan minat siswa, sehingga hasil clustering dapat lebih akurat dan bermanfaat.

3. Melakukan validasi hasil clustering dengan menggunakan metode evaluasi lain seperti Silhouette Score, Davies-Bouldin Index, atau menggunakan teknik cross-validation untuk memastikan bahwa hasil clustering yang diperoleh benar-benar menggambarkan pola data yang ada.
4. Peningkatan infrastruktur teknologi dengan memanfaatkan teknologi big data dan machine learning yang lebih canggih untuk menangani dataset yang lebih besar dan kompleks, sehingga hasil analisis dapat lebih cepat dan akurat.
5. Menambahkan fitur upload CSV/Excel agar pengguna dapat dengan mudah menggunakan sistem tanpa perlu mengubah kodingan. Fitur ini akan mempermudah proses input data, sehingga pengguna bisa langsung mengunggah file data yang diperlukan dan sistem dapat langsung memprosesnya.

## 8. DAFTAR PUSTAKA

- Abdillah, L. A. (2013). Students learning center strategy based on e-learning and blogs. *Seminar Nasional Sains dan Teknologi (SNST) ke- 4 tahun 2013*, pp. F.3.15-20.
- Amir, Z., & Risnawati. (2015). *Psikologi Pembelajaran Matematika*. Yogyakarta: Aswaja Pressindo.
- Arriyanti, E., & Adytia, P. (2020). A Synthesis of Optimal Unknown Number Clustering System and Categorical. *International Journal*, 9(5).
- Badudu, & Zain, S. M. (2010). *Efektifitas Bahasa Indonesia*. Jakarta: Balai Pustaka.
- Dimiyati, & Mudjiono. (2013). *Belajar dan Pembelajaran*. Jakarta: PT. Rineka Cipta.
- DQLab. (2023, Juli 06). *Mengenal Streamlit, Tools Favorit Data Scientist*. Retrieved from [dqlab.id](https://dqlab.id/mengenal-streamlit-tools-favorit-data-scientist): <https://dqlab.id/mengenal-streamlit-tools-favorit-data-scientist>
- Dwilestari, G., Mulyawan, M., & Ali, I. (2021). Analisis Clustering menggunakan K-Medoid pada Data Penduduk Miskin Indonesia. *Jurnal Sistem Informasi Dan Manajemen*, vol. 9, no. 3, pp.282–290.
- Ginting, K. V., Ginting, R. I., & Syaifuddin, M. (2021). Implementasi Data Mining Pengelompokan Minat Belajar Siswa Daring di SMA Negeri 1 Delitua Menggunakan Metode Clustering K-Means. *CyberTech*, 4(8).
- Hakim, M. L. (2020). Pengelompokan Data Transaksi Penjualan menggunakan Algoritma Clustering K-Means pada Toko An Nur. *Universitas Semarang*.
- Hermanto, K., Salim, D., Wu, B., Salim, O. R., & Gunadi, R. B. (2023). Penggunaan Python Untuk Menganalisis Pola Penyebaran Covid-19 Di Masa Pandemi. *Journal of Student Development Information System (JoSDIS)*, Volume 3 Nomor 2.
- Kani, M. K. (2020). *Algoritma dan Pemrograman*. Banten: Universitas Terbuka.
- Munir, R., & Lidyia, L. (2016). *Algoritma dan pemrograman : dalam bahasa pascal, C, dan C++*. Bandung: Informatika Bandung.
- Nugroho, R. (2003). *Prinsip Penerapan Pembelajaran*. Jakarta: Balai Pustaka.
- Prasetyo, E. (2012). *Data Mining - Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Penerbit Andi.
- Purba, W., Siawin, W., Hardih, Nababan, M., Dharshinni, N., & Aisyah, S. (2018). Implementasi Data Mining untuk Pengelompokan dan Prediksi Karyawan yang Berpotensi PHK dengan Algoritma K-Means Clustering. *Jurnal Sistem Informasi Ilmu Komputer Prima*, Vol. 2 No. 1.
- Putri, J. E. (2019). Penerapan Model Pembelajaran LEARNING CYCLE untuk Meningkatkan Hasil Belajar Siswa pada Mata Pelajaran Pendidikan Agama Islam di SMPN 05 Bengkulu Selatan. *Universitas IAIN Bengkulu*.
- Rifai, A. (2022). *19996/3/BAB\_II\_AR*. Retrieved from [repository.uinsu.ac.id](http://repository.uinsu.ac.id): [http://repository.uinsu.ac.id/19996/3/BAB\\_II\\_A\\_R.pdf](http://repository.uinsu.ac.id/19996/3/BAB_II_A_R.pdf)
- Rizky, D. (2019, April 30). *jenis-flowchart-dan-simbol-simbolnya-ef6553c53d73*. Retrieved from [medium.com](https://medium.com/dot-intern/jenis-flowchart-dan-simbol-simbolnya-ef6553c53d73): <https://medium.com/dot-intern/jenis-flowchart-dan-simbol-simbolnya-ef6553c53d73>
- Salsabila, N. (2019). *Klasifikasi Barang menggunakan Metod clustering K-Means dalam Penentuan Prediksi Stok Barang (Studi Kasus : UKM Mar'ah Jilbab Kediri)*. Malang: Universitas Islam Negeri Maulana Malik Ibrahim.
- Sembiring, Y. R., Saifullah, & Winanjaya, R. (2021). Implementasi Data Mining dalam Mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi Menggunakan Algoritma K-Means. *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, 125-132.
- Sembiring, Y. R., Saifullah, & Winanjaya, R. (2021). Implementasi Data Mining Dalam Mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi Menggunakan Algoritma K-Means. *Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, Vol. 2, No. 2, 125-132.
- Setiawan, G. (2004). *Implementasi dalam Birokrasi Pembangunan*. Jakarta: Balai Pustaka.
- Studio, C. (2023, Juli 09). *Visual Studio Code: Pengertian, Fitur, dan Keunggulannya*. Retrieved from [codingstudio.id](https://codingstudio.id): <https://codingstudio.id/blog/visual-studio-code-adalah/>
- Sufajar Butsianto, S. M., & Saepudin. S. Kom, N. (2020). Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means. *Jurnal Nasional Komputasi dan Teknologi Informasi*, 3(1).

- Sujarweni, V. W. (2014). *Metodologi Penelitian*. Yogyakarta: Pustaka Baru Press.
- Suprayogi. (2013). Data Mining Clustering. *Journal of Chemical Information and Modeling*, 1689-1699.
- Suprijono, A. (2015). *Cooperative Learning Teori & Aplikasi PAIKEM*. Yogyakarta: Pustaka Belajar.
- Suryadi, U. T., & Supriatna, Y. (2019). Sistem Clustering Tindak Kejahatan Pencurian Di Wilayah Jawa Barat Menggunakan Algoritma K-Means. *Jurnal Teknologi dan Komunikasi STMIK Subang*, 17.
- Syafitri, I. (2021). *Implementasi Algoritma K-Means Clustering dalam Pengarsipan Dokumen pada BAUK STMIK Widya Cipta Dharma. Jurusan Teknik Informatika*. Samarinda: STMIK Widya Cipta Dharma.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2014). *Introduction to Data Mining*. Pearson.
- Usman, N. (2002). *Konteks Implementasi Berbasis Kurikulum*. Bandung: Grasindo.
- Wahyudi, M., Masitha, Saragih, R., & Solikhun. (2020). *Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*. Medan: Yayasan Kita Menulis.
- Wibowo, T. (2018, 08 10). *naskah\_publicasi\_L200144019\_INF.pdf*. Retrieved from eprints.ums.ac.id: [https://eprints.ums.ac.id/65663/3/naskah\\_publicasi\\_L200144019\\_INF.pdf](https://eprints.ums.ac.id/65663/3/naskah_publicasi_L200144019_INF.pdf)
- xerma. (2015). *kumpulan artikel tentang ilmu komputer, internet dan juga pelajaran sekolah dan artikel lainnya*. Retrieved from Kumpulan Artikel News: <http://xerma.blogspot.com/2015/01/pengertian-data-mining-apa-itu-data.html>
- Yulianto, H. S. (2022, Juni 29). *Pengertian Clustering, Metode, dan Syarat-syaratnya*. Retrieved from bola.com: <https://www.bola.com/ragam/read/4997478/pengertian-clustering-metode-dan-syarat-syaratnya?page=3>