PERBANDINGAN ALGORITMA K-MEANS CLUSTERING DAN FUZZY TSUKAMOTO UNTUK MENGELOMPOKKAN DATA SISWA SMP NU MEDAN BERDASARKAN PRESTASI AKADEMIK

PROPOSAL SKRIPSI

DISUSUN OLEH

ABDUL RASYID PAHLEVI
2109020064

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

MEDAN

2025

Perbandingan Algoritma K-Means Clustering dan Fuzzy Tsukamoto Untuk Mengelompokkan Data Siswa SMP NU Medan Berdasarkan Prestasi Akademik

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) dalam Program Studi Teknologi Informasi pada Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara

Abdul Rasyid Pahlevi
NPM. 2109020064

PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025

LEMBAR PENGESAHAN

Judul Skripsi	:	PERBANDINGAN ALGORITMA K-MEANS CLUSTERING DAN FUZZY TSUKAMOTO UNTUK MENGELOMPOKKAN DATA SISWA SMP NU MEDAN BERDASARKAN PRESTASI AKADEMIK
Nama Mahasiswa	:	ABDUL RASYID PAHLEVI
NPM	:	2109020064
Program Studi	:	TEKNOLOGI INFORMASI

Menyetujui
Komisi Pembimbing

(Amrullah, S.Kom., M.Kom)
NIDN. 0125118604

Ketua Program Studi

(Fatmasari Hutagalung, S,Kom. M.Kom.)
NIDN. 0117088902

Dekan

(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

Perbandingan Algoritma K-Means Clustering dan Fuzzy Tsukamoto Untuk Mengelompokkan Data Siswa SMP NU Medan Berdasarkan Prestasi Akademik

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, Agustus 2025

Yang membuat pernyataan

Abdul Rasyid Pahlevi

NPM. 2109020060

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama	: Abdul Rasyid Pahlevi
NPM	2109020064
Program Studi	: Teknologi Informasi
Karya Ilmiah	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif *(Non-Exclusive Royalty free Right)* atas penelitian skripsi saya yang berjudul:

Perbandingan Algoritma K-Means Clustering dan Fuzzy Tsukamoto Untuk Mengelompokkan Data Siswa SMP NU Medan Berdasarkan Prestasi Akademik

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, Agustus 2025

Yang membuat pernyataan

Abdul Rasyid Pahlevi

NPM. 2109020064

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap	: Abdul Rasyid Pahlevi
Tempat dan Tanggal Lahir	: Pontianak, 04 Oktober 2003
Alamat Rumah	: Desa Sungai Durian
Telepon/Faks/HP	: 081262130813
E-mail	: rasyidabdul453@gmail.com
Instansi Tempat Kerja	: -
Alamat Kantor	: -

DATA PENDIDIKAN

SD	: SD Negeri 01 Gunung Tua	TAMAT: 2015
SMP	: MTSN 2 Padang Bolak	TAMAT: 2018
SMA	: SMAN 1 Padang Bolak	TAMAT: 2021

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadirat Tuhan Yang Maha Esa, penulis akhirnya dapat menyelesaikan penyusunan skripsi yang berjudul "Optimasi Model Prediksi Hasil Belajar Siswa dalam Pembelajaran IPA Berbasis Media Sosial Menggunakan XGBoost dan Bayesian Optimization" sebagai salah satu bentuk tanggung jawab akademik dan syarat dalam menyelesaikan studi di Program Studi Teknologi Informasi, Fakultas Ilmu computer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara. Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Fatma Sari Hutagalung S.Kom., M.Kom. Ketua Program Studi Teknologi Informasi
4. Bapak Mhd. Basri, S.Si, M.Kom Sekretaris Program Studi Teknologi Informasi
5. Bapak Amrullah, S.Kom, M.Kom. Pembimbing Skripsi
6. Ayah tercinta H.Burhan Harahap SH.MH., semangat dan doa yang telah engkau berikan selama hidupku akan selalu jadi semangat disetiap langkahku. Terima kasih atas semua dedikasi yang telah diberikan selama ini dan selalu menjadi garis terdepan untuk penulis.
7. Ibunda tercinta saya, Deswati Siregar yang selalu mendoakan saya setiap hari, memberi semangat, dan menjadi sumber kekuatan dalam setiap langkah penulis.

8. Abang dan kakak penulis yang selalu memberikan semangat dan dukungan, baik secara langsung maupun tidak langsung. Terima kasih atas dorongan yang membuat penulis menjadi semangat dan tidak menyerah dalam melewati proses ini.
9. Penulis mengucapkan terima kasih kepada Inka Minta Ito Siregar, A.Md.Stat., yang selalu menjadi sumber kekuatan, motivasi, dan inspirasi. Dukungan, doa, serta kesabaranmu adalah salah satu alasan terbesar penulis mampu menyelesaikan ini hingga akhir
10. Semua pihak yang terlibat langsung ataupun tidak langsung yang tidak dapat penulis ucapkan satu-persatu yang telah membantu penyelesaian skripsi ini.

PERBANDINGAN ALGORITMA K-MEANS CLUSTERING DAN FUZZY TSUKAMOTO UNTUK MENGELOMPOKKAN DATA SISWA SMP NU MEDAN BERDASARKAN PRESTASI AKADEMIK

ABSTRAK

Penelitian ini bertujuan untuk membandingkan kinerja algoritma K-Means Clustering dan Fuzzy Tsukamoto dalam mengelompokkan data siswa SMP NU Medan berdasarkan prestasi akademik. Latar belakang penelitian ini didasarkan pada pentingnya analisis data pendidikan sebagai dasar pengambilan keputusan yang tepat dalam meningkatkan kualitas pembelajaran. Data yang digunakan berasal dari nilai akademik siswa pada beberapa mata pelajaran inti. Metode K-Means Clustering digunakan untuk melakukan pengelompokan secara tegas (hard clustering), sedangkan Fuzzy Tsukamoto digunakan untuk mengelompokkan dengan pendekatan logika fuzzy (soft clustering). Hasil penelitian menunjukkan bahwa algoritma K-Means mampu menghasilkan pengelompokan yang sederhana dan mudah diinterpretasikan, namun cenderung kurang fleksibel terhadap data yang bersifat ambigu. Sebaliknya, metode Fuzzy Tsukamoto memberikan hasil pengelompokan yang lebih adaptif dengan mempertimbangkan derajat keanggotaan setiap data, sehingga lebih sesuai dalam merepresentasikan variasi prestasi akademik siswa. Kesimpulan dari penelitian ini adalah Fuzzy Tsukamoto lebih unggul dalam memberikan analisis yang komprehensif terhadap data prestasi akademik, sedangkan K-Means lebih unggul dari sisi kemudahan implementasi.

Kata Kunci: Data Mining, K-Means, Fuzzy Tsukamoto, Clustering, Prestasi Akademi

A Comparative Study of K-Means Clustering and Fuzzy Tsukamoto Algorithms for Classifying Students' Academic Achievement at SMP NU Medan

ABSTRACT

This research aims to compare the performance of the K-Means Clustering algorithm and the Fuzzy Tsukamoto method in grouping students of SMP NU Medan based on their academic achievement. The background of this study lies in the importance of educational data analysis as a foundation for decision-making to improve learning quality. The dataset consists of students' grades from core subjects collected over four semesters. The K-Means algorithm was applied to perform hard clustering, where each student is assigned to a single cluster, while the Fuzzy Tsukamoto method was applied to perform soft classification using fuzzy logic rules. The results indicate that K-Means produces clusters that are relatively proportional and easy to interpret, but less flexible in handling borderline cases. In contrast, Fuzzy Tsukamoto provides more adaptive results by considering degrees of membership, which allows for a more comprehensive representation of students' academic performance. The conclusion of this study highlights that Fuzzy Tsukamoto is more suitable for absolute classification based on score thresholds, while K-Means is better suited for relative grouping within a class context.

Keywords: Data Mining, K-Means, Fuzzy Tsukamoto, Clustering, Academic Performance.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Pendidikan merupakan salah satu aspek fundamental dalam pembangunan sumber daya manusia yang berkualitas. Dalam konteks pendidikan formal, prestasi akademik siswa menjadi indikator utama untuk mengukur keberhasilan proses pembelajaran dan kualitas pendidikan di suatu institusi. Prestasi akademik tidak hanya mencerminkan kemampuan kognitif siswa, tetapi juga menjadi dasar bagi pengambilan keputusan strategis dalam pengelolaan Pendidikan (Rezky et al., 2019).

SMP NU (Nahdlatul Ulama) Medan sebagai salah satu institusi pendidikan menengah pertama di kota Medan memiliki komitmen untuk terus meningkatkan kualitas pendidikan. Dengan jumlah siswa yang terus bertambah setiap tahunnya, sekolah menghadapi tantangan dalam mengelola dan menganalisis data prestasi akademik siswa secara efektif. Data prestasi akademik yang terkumpul dari berbagai mata pelajaran dan periode pembelajaran memerlukan pendekatan analisis yang sistematis untuk dapat memberikan insight yang berguna bagi pengambilan keputusan pendidikan.

Dalam era digital saat ini, pemanfaatan teknologi informasi dan teknik data mining menjadi solusi yang tepat untuk menganalisis data dalam jumlah besar(Papakyriakou & Barbounakis, 2022). Salah satu teknik data mining yang populer dan efektif untuk pengelompokan data adalah clustering(Gupta et al., 2015). Clustering adalah teknik unsupervised learning yang bertujuan untuk

mengelompokkan data ke dalam beberapa cluster berdasarkan kesamaan karakteristik tertentu(Saligkaras & Papageorgiou, n.d.).

Algoritma K-Means merupakan salah satu algoritma clustering yang paling banyak digunakan karena kesederhanaannya dan efisiensinya dalam menangani dataset berukuran besar(Priya P & Souza, n.d.). Algoritma ini bekerja dengan cara membagi data ke dalam k cluster, dimana setiap data point akan menjadi anggota dari cluster yang memiliki centroid terdekat(Adzika & Djagba, 2024).

Pengelompokan siswa berdasarkan prestasi akademik memiliki manfaat yang signifikan bagi pengelolaan pendidikan. Pertama, pengelompokan ini dapat membantu guru dan pihak sekolah dalam mengidentifikasi siswa yang memerlukan perhatian khusus, baik siswa yang berprestasi tinggi maupun siswa yang mengalami kesulitan belajar. Kedua, hasil pengelompokan dapat menjadi dasar untuk merancang strategi pembelajaran yang lebih personal dan sesuai dengan kebutuhan setiap kelompok siswa. Ketiga, analisis ini dapat membantu dalam alokasi sumber daya pendidikan yang lebih efisien dan efektif.

Penelitian yang dilakukan oleh Fenty Eka M. Agustin dan kawan-kawan dengan judul "Implementasi Algoritma K-Means Untuk Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus: Smp Negeri 101 Jakarta)" menunjukkan bahwa algoritma K-Means dapat diimplementasikan untuk membantu pengelompokkan kemampuan siswa terhadap mata pelajaran Ujian Nasional. Namun, penelitian ini terbatas pada data ujian nasional saja dan tidak mempertimbangkan nilai mata pelajaran secara individual(Eka Agustin et al., 2015).

Hendro Priyatman dan kawan dalam penelitiannya " Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa" mengaplikasikan K-Means untuk mengelompokkan mahasiswa berdasarkan IPK. Hasil penelitian menunjukkan bahwa pada kasus ini implementasi algoritma k-means dalam data mining sudah berhasil, dan bisa menampilkan informasi prediksi kelulusan mahasiswa, namun untuk melihat tingkat kemampuan real k-means clustering dalam memprediksi waktu kelulusan tergantung pada mahasiswa itu sendiri(Priyatman et al., 2019).

Studi yang dilakukan Diwa Oktario Dacwanda dan kawan-kawan dengan judul " Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan " penelitian ini menyimpulkan bahawa Penerapan algoritma k-Means dalam pengelompokan nilai pengetahuan dan keterampilan siswa menghasilkan tiga cluster, yaitu pintar, sedang, dan cukup, dengan hasil yang menunjukkan bahwa pengelompokan berdasarkan kedua nilai tersebut lebih akurat dibandingkan hanya berdasarkan nilai pengetahuan, meskipun perpindahan siswa antar cluster hanya sebesar 10,15%. Hasil ini bermanfaat bagi sekolah dalam merancang sistem pengajaran yang seimbang antara pengetahuan dan keterampilan, serta membuka peluang pengembangan penelitian lebih lanjut dengan mempertimbangkan nilai non-akademis seperti ekstrakurikuler dan membandingkan metode k-Means dengan metode lain seperti Fuzzy C-Means(Oktario Dacwanda & Nataliani, 2021).

Dari penelitian-penelitian terdahulu, dapat diidentifikasi beberapa gap dan peluang pengembangan, seperti kurangnya perhatian terhadap konteks spesifik institusi, penggunaan variabel yang terbatas hanya pada nilai rata-rata atau IPK

sedangkan analisis per mata pelajaran dapat memberikan insight yang lebih detail. Algoritma K-Means populer karena kesederhanaan dan efisiensinya. Namun, K-Means bekerja pada batasan keanggotaan tegas (*hard partition*) setiap siswa hanya menjadi anggota satu klaster. Pada kenyataannya, prestasi akademik sering bersifat kontinu dan mengandung ketidakpastian, misalnya siswa yang berada di ambang antara kategori 'sedang' dan 'tinggi'. Untuk itu, penelitian ini menambahkan metode Fuzzy Tsukamoto sebagai pembanding. Metode ini memodelkan derajat keanggotaan (membership) sehingga setiap siswa dapat memiliki derajat rendah/sedang/tinggi sekaligus, lalu dilakukan defuzzifikasi untuk memperoleh skor dan kategori akhir. Perbandingan kedua pendekatan diharapkan memberi hasil yang lebih komprehensif dan interpretatif bagi guru dan pengelola sekolah. Oleh karena itu, penelitian ini berupaya mengisi gap tersebut dengan fokus pada konteks spesifik SMP NU Medan, menggunakan data nilai per mata pelajaran, melakukan evaluasi clustering yang komprehensif, serta memberikan rekomendasi praktis yang dapat diterapkan untuk meningkatkan kualitas pengelolaan pendidikan di institusi tersebut melalui pemanfaatan teknologi data mining.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana menerapkan algoritma K-Means untuk mengelompokkan data siswa SMP NU Medan berdasarkan prestasi akademik?
2. Bagaimana menerapkan metode Fuzzy Tsukamoto untuk mengelompokkan/ mengklasifikasikan tingkat prestasi akademik siswa?

1.3 Pembatasan Masalah

Untuk memfokuskan penelitian dan menghasilkan analisis yang mendalam, penelitian ini memiliki batasan-batasan sebagai berikut:

1. Data yang digunakan adalah data prestasi akademik siswa SMP NU Medan untuk mata pelajaran inti yaitu PAI, PPKN, Matematika, Bahasa Indonesia, Bahasa Inggris, IPA (Ilmu Pengetahuan Alam), dan IPS (Ilmu Pengetahuan Sosial), Senibudaya, PJOK, Prakarya, IQRA dan TIK.
2. Periode data yang dianalisis adalah data nilai siswa dari semester I - IV
3. Metode yang digunakan: K-Means (clustering) dan Fuzzy Tsukamoto (logika fuzzy) sebagai pembanding
4. Jumlah cluster yang dibentuk adalah 3 cluster, yaitu siswa dengan prestasi tinggi, sedang, dan rendah.
5. Penelitian ini tidak membahas faktor-faktor eksternal yang mempengaruhi prestasi akademik siswa seperti kondisi sosial ekonomi, motivasi belajar, atau metode pembelajaran.
6. Implementasi algoritma menggunakan bahasa pemrograman Python dengan library scikit-learn.

1.4 Tujuan Penelitian

Penelitian ini memiliki tujuan umum dan tujuan khusus sebagai berikut:

Tujuan Umum :

Menerapkan algoritma K-Means Clustering dan Fuzzy Tsukamoto untuk mengelompokkan data siswa SMP NU Medan berdasarkan prestasi akademik guna mendukung pengambilan keputusan dalam pengelolaan pendidikan.

Tujuan Khusus :

1. Mengimplementasikan algoritma K-Means Clustering dan Fuzzy Tsukamoto untuk mengelompokkan data prestasi akademik siswa SMP NU Medan.
2. Mengidentifikasi karakteristik dan pola prestasi akademik dari setiap cluster yang terbentuk.
3. Memberikan rekomendasi strategi pembelajaran berdasarkan hasil pengelompokan siswa.
4. Menghasilkan sistem atau tools yang dapat digunakan oleh pihak sekolah untuk melakukan pengelompokan siswa secara otomatis.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat bagi berbagai pihak:

Bagi SMP NU Medan:

1. Membantu pihak sekolah dalam mengidentifikasi kelompok siswa berdasarkan prestasi akademik.
2. Menyediakan dasar untuk merancang strategi pembelajaran yang lebih personal dan efektif.
3. Memfasilitasi alokasi sumber daya pendidikan yang lebih optimal.
4. Mendukung sistem evaluasi dan monitoring prestasi siswa.
5. Membantu guru dalam memahami karakteristik siswa di kelasnya.

6. Memberikan guidance untuk merancang metode pembelajaran yang sesuai dengan kemampuan siswa.
7. Memfasilitasi pemberian perhatian khusus kepada siswa yang memerlukan.

Bagi Peneliti Lain:

1. Menyediakan referensi untuk penelitian sejenis di bidang educational data mining.
2. Memberikan contoh implementasi algoritma K-Means dalam konteks pendidikan.

BAB II

KAJIAN PUSTAKAN

2.1 Data Mining

Data mining adalah proses ekstraksi pola atau informasi yang berguna dari kumpulan data yang besar. Dalam konteks pendidikan, data mining telah menjadi alat yang powerful untuk menganalisis berbagai aspek proses pembelajaran dan menghasilkan insight yang dapat meningkatkan kualitas pendidikan(Yuli Mardi, n.d.2020).

2.1.1 Defenisi Data Mining

Menurut Han dan Kamber (2012), data mining didefinisikan sebagai proses penemuan pola yang menarik dan pengetahuan dari data dalam jumlah besar. Data mining melibatkan integrasi teknik dari beberapa bidang seperti statistik, machine learning, database systems, dan visualization(Putra & Wadisman, 2018). (Karlina & Nurdiawan, 2023)mendefinisikan data mining sebagai langkah analisis dalam proses Knowledge Discovery in Databases (KDD). Proses KDD meliputi tahapan: selection, preprocessing, transformation, data mining, dan interpretation/evaluation.

2.1.2 Tahapan Data Mining

Proses data mining umumnya mengikuti metodologi CRISP-DM (Cross-Industry Standard Process for Data Mining) yang terdiri dari enam tahap(Ariesanto Akhmad, 2020):

1. Business Understanding: Memahami objektif dan kebutuhan dari perspektif bisnis, kemudian mendefinisikan masalah data mining.
2. Data Understanding: Mengumpulkan data awal dan melakukan eksplorasi untuk memahami karakteristik data.
3. Data Preparation: Mempersiapkan dataset final dari data mentah, termasuk cleaning, integration, selection, dan transformation.
4. Modeling: Memilih dan menerapkan teknik modeling yang appropriate, serta kalibrasi parameter untuk nilai optimal.
5. Evaluation: Mengevaluasi model secara menyeluruh dan review tahapan untuk memastikan model mencapai objektif bisnis.
6. Deployment: Mengatur penggunaan model dalam lingkungan produksi.

2.1.3 Teknik Data Mining

Data mining memiliki berbagai teknik yang dapat dikategorikan berdasarkan tujuan analisisnya(Rizal, n.d.):

Supervised Learning:

1. Classification: Memprediksi kategori atau kelas dari data baru
2. Regression: Memprediksi nilai numerik kontinu.

Unsupervised Learning:

1. Clustering: Mengelompokkan data berdasarkan kesamaan
2. Association Rules: Menemukan hubungan antar item dalam dataset

Semi-supervised Learning:

1. Kombinasi antara supervised dan unsupervised learning

2.1.4 Data Mining Dalam Pendidikan

Educational Data Mining (EDM) adalah area penelitian yang berkembang pesat yang menerapkan teknik data mining pada data pendidikan(Anastassia et al., 2024). Menurut (Ardianti et al., 2024), EDM bertujuan untuk mengembangkan metode eksplorasi data unik yang berasal dari lingkungan pendidikan dan menggunakan metode tersebut untuk lebih memahami siswa dan setting pembelajaran.

Aplikasi EDM meliputi:

1. Analisis performa akademik siswa
2. Personalisasi pembelajaran
3. Prediksi dropout dan kelulusan
4. Rekomendasi kurikulum
5. Evaluasi efektivitas pembelajaran
6. Deteksi kecurangan akademik

2.2 Clustering

Clustering adalah teknik unsupervised learning yang bertujuan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kesamaan karakteristik. Dalam clustering, tidak ada target variable yang ditentukan sebelumnya, sehingga algoritma harus menemukan struktur tersembunyi dalam data(Adawiyah et al., 2024).

Klasterisasi merupakan metode yang sangat berperan penting dalam kegiatan penelitian untuk menyelesaikan berbagai permasalahan di sejumlah bidang seperti arkeologi, psikiatri, teknik, dan kedokteran. Klaster atau kelompok dibentuk dari sekumpulan titik data yang memiliki kemiripan satu sama lain, namun

berbeda dengan titik data dari klaster lainnya. Teknik klasterisasi juga umum dimanfaatkan di berbagai disiplin ilmu, seperti analisis jejaring sosial, deteksi kriminalitas, hingga pengembangan perangkat lunak, karena kemampuannya dalam mengenali pola dan mengelompokkan data berdasarkan kesamaan karakteristik. Keunggulan klasterisasi terletak pada fleksibilitas penerapannya di bidang pengenalan pola, machine learning, pemrosesan citra, dan statistik. Tujuan utamanya adalah untuk mengelompokkan data dengan pola yang serupa ke dalam kategori-kategori yang terpisah(Marcelina et al., 2023).

Algoritma adalah metode perhitungan yang digunakan komputer untuk menyelesaikan suatu persoalan(Nurul Fatma Dewi Mardianto & Yahfizham Yahfizham, 2024). Ia juga merupakan salah satu fondasi utama dalam bidang ilmu komputer. Untuk memperkirakan banyaknya langkah yang diperlukan, digunakan pendekatan asimtotik. Asimtotik membantu melihat efisiensi algoritma dalam skala besar. Pendekatan ini mengabaikan nilai konstanta dan komponen yang lebih kecil. Sehingga hanya fokus pada pertumbuhan fungsi utama dari algoritma tersebut.

Inti dari algoritma klasterisasi adalah memperhatikan titik pusat data sebagai pusat klaster yang membedakan antar kelompok. Tujuan utamanya adalah membentuk struktur hubungan hierarki antar data dalam proses pengelompokan. Algoritma klasterisasi yang paling dikenal sepanjang waktu antara lain K-Means dan K-Medoids(Nasution & Eka, 2018). Algoritma ini bertujuan untuk mengelompokkan objek-objek yang memiliki kemiripan tinggi ke dalam satu klaster. Sementara itu, objek yang memiliki perbedaan signifikan akan dimasukkan ke dalam klaster yang berbeda. Berkat kemampuannya dalam mencari solusi

optimal, algoritma ini menjadi andalan dalam menyelesaikan masalah pengelompokan.

Menurut Efran Fernando dan kawan-kawannya, klasterisasi adalah proses memisahkan kumpulan objek data ke dalam beberapa bagian atau subset. Setiap subset disebut sebagai klaster, di mana objek-objek di dalamnya memiliki kemiripan yang tinggi(Fernando Ade Pratama & Jumadi, n.d.). Sebaliknya, objek dari klaster yang berbeda menunjukkan perbedaan yang signifikan. Klasterisasi juga dapat dilihat sebagai persoalan optimasi dalam mencari pembagian data terbaik. Sementara itu, Sabrina Aulia Rahmah menyatakan bahwa klasterisasi merupakan proses pengelompokan data berdasarkan kesamaan. Kesamaan di sini berarti setiap objek dalam satu grup memiliki ciri-ciri atau hubungan yang saling berkaitan(Sabrina Aulia Rahmah, n.d.).

2.2.1 Jenis Clustering

Berdasarkan metodologi yang digunakan, clustering dapat diklasifikasikan menjadi beberapa kategori(Elisna et al., n.d.):

1. Partitional Clustering
    a. Membagi data ke dalam k partisi non-overlapping
    b. Setiap objek hanya menjadi anggota satu cluster
2. Hierarchical Clustering
    a. Membentuk hirarki cluster dalam bentuk tree structure
3. Density-based Clustering
    a. Mengelompokkan objek berdasarkan kepadatan area
    b. Dapat menemukan cluster dengan bentuk arbitrary

4. Grid-based Clustering
    a. Membagi ruang data ke dalam grid cells
    b. Clustering dilakukan pada grid cells

5. Model-based Clustering
    a. Mengasumsikan model matematika untuk setiap cluster
    b. Menggunakan statistical approach

2.3.2 Aplikasi Clustering dalam Pendidikan

Dalam konteks pendidikan, clustering memiliki berbagai aplikasi:

1. Student Grouping: Mengelompokkan siswa berdasarkan kemampuan, gaya belajar, atau karakteristik lain
2. Curriculum Analysis: Mengelompokkan mata pelajaran berdasarkan tingkat kesulitan atau korelasi
3. Learning Path Recommendation: Mengidentifikasi jalur pembelajaran yang optimal
4. Teacher Performance Analysis: Mengelompokkan guru berdasarkan efektivitas mengajar
5. Resource Allocation: Optimalisasi distribusi sumber daya Pendidikan

2.3 Algoritma K-Means

K-Means adalah salah satu algoritma clustering yang paling populer dan banyak digunakan(Putra Eriansya & Syafrullah, 2018). Algoritma ini pertama kali diperkenalkan oleh MacQueen (1967) dan telah menjadi standar untuk many clustering applications karena kesederhanaan dan efisiensinya.

Klasterisasi dengan algoritma K-Means adalah metode untuk membagi data ke dalam kelompok yang memiliki kemiripan(Zuhal, 2022). Setiap kelompok atau klaster dibentuk berdasarkan kesamaan karakteristik antar data. Salah satu keunggulan K-Means adalah kemampuannya dalam memproses data berukuran besar dengan cepat dan efisien. Algoritma ini sangat cocok digunakan untuk analisis data yang memerlukan segmentasi secara otomatis. Langkah awal dalam K-Means adalah menentukan titik pusat (centroid) secara acak sebagai dasar pengelompokan.

Langkah awal dalam algoritma K-Means dimulai dengan pemilihan titik pusat (centroid) secara acak. Proses pengelompokan dalam algoritma ini dilakukan dengan cara meminimalkan tingkat kesalahan dalam penempatan data ke dalam klaster yang tepat(Putri et al., 2024). Karena sifatnya yang sederhana, praktis, dan sangat efisien, K-Means telah berhasil digunakan di berbagai bidang, seperti pengelompokan dokumen, segmentasi pasar, pengolahan citra, klasifikasi nasabah, hingga ekstraksi fitur. Secara umum, metode klasterisasi terbagi menjadi dua kategori utama, yaitu klasterisasi hierarki dan klasterisasi partisi.

Dalam bidang pendidikan, algoritma kmeans digunakan untuk mengelompokkan siswa berdasarkan kesamaan karakter pribadi dan gaya belajar, khususnya untuk membantu siswa yang kurang percaya diri karena lambat memahami materi. Melalui klasterisasi, masing-masing siswa dikelompokkan berdasarkan tingkat kemampuan dan cara belajarnya(Sapriadi et al., 2023).

Algoritma K-Means memiliki kemampuan untuk mengelompokkan data berukuran besar, termasuk data multiview yang berasal dari berbagai tabel atau dataset yang berbeda(Nur et al., 2018). Proses pengelompokan dilakukan secara

simultan pada setiap tabel, sehingga membentuk pendekatan multiview clustering. Teknik ini dikenal sebagai metode pengelompokan berbasis centroid, di mana titik pusat awal dijadikan representasi dari jumlah klaster yang ditentukan. Kualitas hasil pengelompokan sangat bergantung pada nilai centroid akhir yang bersifat konstan dan menjadi acuan pembentukan klaster yang optimal.

Tujuan dari algoritma K-Means adalah mengelompokkan sejumlah objek ke dalam k klaster, di mana nilai k telah ditentukan terlebih dahulu. Setiap klaster memiliki centroid atau titik pusat yang dihitung sebagai rata-rata dari seluruh objek yang berada dalam klaster tersebut(Yunita, 2018). Dalam proses pengelompokannya, algoritma ini menggunakan jarak Euclidean untuk mengukur tingkat kemiripan antar objek data. Adapun tahapan dalam algoritma K-Means dilakukan secara sistematis sebagai berikut:

1. *Preprocessing*

   Sebelum menerapkan algoritma K-Means, tahap awal yang harus dilakukan adalah preprocessing data untuk memperoleh hasil klasterisasi yang optimal. Proses preprocessing meliputi beberapa langkah seperti seleksi data, transformasi, dan normalisasi, tergantung pada kebutuhan dan karakteristik algoritma yang digunakan. Pada penelitian ini, proses normalisasi dilakukan dengan menggunakan metode MinMax, yaitu metode yang berfungsi untuk menyetarakan skala antar data, terutama antara nilai yang besar dan nilai yang kecil. Hal ini penting karena perbedaan skala dapat menyebabkan penyimpangan dalam proses klasifikasi. Dengan metode MinMax, semua nilai akan disesuaikan dalam rentang yang sama agar perhitungan lebih

akurat. Adapun proses normalisasi MinMax dinyatakan dalam persamaan sebagai berikut:

$$x = \frac{x - nilai\ min}{nilai\ \max - nilai\ min}$$

Dimana: x = data perkolom

nilai min = nilai terkecil perkolom

nilai max = nilai terbesar perkolom

2. Menentukan jumlah klaster (nilai K) dan pusat klaster atau centroid awal. Langkah selanjutnya adalah menentukan jumlah klaster (K) serta menetapkan titik pusat awal (centroid) untuk setiap klaster. Jumlah klaster umumnya ditentukan berdasarkan jumlah kategori atau segmen data yang diinginkan sesuai dengan tujuan analisis. Dalam algoritma K-Means, centroid awal dipilih secara acak pada setiap klaster sebagai titik awal dalam proses perhitungan jarak Euclidean. Untuk memperoleh nilai K yang optimal, dapat digunakan metode *Elbow*, yaitu metode yang menentukan jumlah klaster terbaik berdasarkan nilai *Within Sum of Square Error* (WSSE). Nilai WSSE menunjukkan jumlah total kuadrat jarak antara data dengan centroid-nya. Titik "tekuk" (*elbow*) pada grafik WSSE terhadap nilai K menunjukkan jumlah klaster yang ideal. Adapun proses perhitungan WSSE dapat dijelaskan melalui persamaan berikut:

$$\sum_{i}^{n} jarak\ (p_{1,}c_i)^2$$

Dimana :

Pi = data ke i

Ci = centroid ke i

3. Menghitung jarak Euclidean

Pada tahap ini, dilakukan perhitungan jarak antara setiap data dengan masing-masing centroid menggunakan rumus *Euclidean Distance*. Jarak ini digunakan untuk menentukan seberapa dekat suatu data terhadap pusat klaster yang telah ditentukan. Nilai jarak terkecil dari hasil perhitungan akan dianggap sebagai jarak terdekat, sehingga data tersebut akan ditempatkan ke dalam klaster dengan centroid terdekat. Proses ini penting dalam menentukan pembentukan awal kelompok yang seragam. Adapun rumus untuk menghitung jarak Euclidean ditunjukkan pada persamaan berikut:

$$d(x, c_i) \sqrt{\sum_{j=1}^{m} (x_j - c_i j)^2}$$

Dimana :

d = jarak pada x dan c

x = nilai centroid awal

c = nilai dari baris N

m = jumlah atribut data

i & j = jumlah iterasi

Nilai terkecil dari hasil rumusan di atas yang dipilih sebagai jarak terdekat ke nilai centroid setiap klister

4. Menghitung rata-rata nilai objek data di setiap klaster sebagai nilai centroid yang baru.

   Setelah proses pengelompokan awal berdasarkan jarak terdekat selesai, langkah selanjutnya adalah menghitung nilai rata-rata dari seluruh objek data dalam setiap klaster. Nilai rata-rata ini akan menjadi centroid baru yang menggantikan posisi centroid awal. Centroid baru tersebut kemudian digunakan dalam perhitungan jarak Euclidean pada iterasi selanjutnya. Proses ini bertujuan untuk memperbarui pusat klaster agar semakin representatif terhadap data yang tergabung di dalamnya.

5. Proses perhitungan dilanjutkan dengan melakukan iterasi secara berulang pada tahap penentuan centroid baru dan perhitungan jarak Euclidean, hingga posisi centroid tidak lagi berubah atau telah mencapai nilai yang konstan. Iterasi ini menunjukkan bahwa proses klasterisasi telah mencapai titik konvergen. Untuk mengevaluasi kualitas hasil klasterisasi, dilakukan pengujian menggunakan Sum of Square Error (SSE), yaitu metrik yang mengukur seberapa dekat data dalam suatu klaster terhadap centroid-nya. Nilai SSE yang semakin kecil menunjukkan bahwa hasil pengelompokan semakin baik. Adapun rumus untuk menghitung nilai SSE dapat dijelaskan melalui persamaan berikut:

$$SSE = \sum_{k=1}^{k} (x_i - c_k)^2$$

Nilai *Sum of Square Error* (SSE) yang kecil hingga mendekati nol menunjukkan bahwa kualitas klasterisasi yang dihasilkan sangat baik. Hal ini mengindikasikan

bahwa objek-objek dalam setiap klaster memiliki jarak yang relatif dekat terhadap centroid-nya. Sebaliknya, jika nilai SSE yang dihasilkan tinggi, maka hal tersebut menunjukkan bahwa hasil pengelompokan kurang optimal atau tidak akurat(Liu et al., 2024).

2.4 Algoritma Fuzzy Tsukamoto

Metode Tsukamoto adalah salah satu teknik dalam sistem logika fuzzy yang digunakan untuk menangani masalah dengan ketidakpastian dan ketidakjelasan data. Dalam metode ini, setiap aturan dalam sistem fuzzy direpresentasikan dalam bentuk aturan "IF-THEN" yang memiliki nilai keanggotaan fuzzy yang terkait dengan output. Metode Tsukamoto merupakan perluasan dari penalaran monoton . Metode fuzzy tsukamoto merupakan metode yang digunakan untuk membantu dalam pemberian rekomendasi secara cepat, tepat, dan akurat.(Pujiarso Nugroho et al., 2019)

Istilah yang digunakan dalam fuzzy adalah sebagai berikut:

a. Derajat Keanggotaan (Degree of Membership)

   Fungsi dari derajat keanggotaan adalah untuk memberikan bobot pada suatu input yang telah diberikan, sehingga input tadi dapat dinyatakan dengan nilai.

   Contoh : Bilangan dari 0 – 1

b. Variabel Fuzzy

   Variabel Fuzzy merupakan variabel yang hendak dibahas dalam suatu aplikasi fuzzy.

   Contoh : Fuzzy yang membahas tentang UMUR

c. Domain (Scope)

Domain himpunan fuzzy adalah keseluruhan nilai yang diijinkan dalam semesta pembicaraan dan boleh dioperasikan dalam suatu himpunan fuzzy. Seperti halnya semesta pembicaraan, domain merupakan himpunan bilangan riil yang senantiasa naik (bertambah) secara monoton dari kiri ke kanan. Nilai domain dapat berupa bilangan positif maupun negatif

d. Label

Label adalah kata-kata untuk memberikan suatu keterangan pada domain.

Contoh : MUDA

PARUBAYA

TUA

e. Fungsi Keanggotaan

Fungsi keanggotaan (membership function) adalah suatu kurva yang menunjukan pemetaan titik input data ke dalam nilai keanggotaannya (sering juga disebut dengan derajat keanggotaan) yang memiliki interval antara 0 sampai 1**.**

Langkah penggunaan fuzzy logic adalah sebagai berikut(Ayu et al., 2018):

1. Mendefinisikan obyektif dan kriteria kontrol
2. Menentukan hubungan antara input dan output serta memilih jumlah minimum variabel input pada mesin fuzzy logic
3. Dengan menggunakan struktur berbasis aturan dari fuzzy logic, menjabarkan permasalahan kontrol ke dalam aturan IF X AND Y THEN Z yang mendefinisikan respon output aplikasi yang diinginkan untuk kondisi input aplikasi diberikan. Jumlah dan kompleksitas dari rules bergantung pada jumlah parameter input yang diproses dan jumlah variabel fuzzy yang bekerjasama dengan tiap–tiap parameter. Jika mungkin, gunakan setidaknya satu variabel

dan turunan waktunya. Walaupun mungkin untuk menggunakan sebuah parameter tunggal yang error saat itu juga tanpa mengetahui rata–rata perubahannya, hal ini melumpuhkan kemampuan aplikasi untuk meminimalisasi keterlampauan untuk sebuah tingkat input.

4. Membuat fungsi keanggotaan yang menjelaskan nilai input atau output yang digunakan didalam rules.
5. Mengetest aplikasi, mengevaluasi hasil, mengatur rules dan fungsi keanggotaan, dan retest sampai hasil yang memuaskan didapat.

BAB III

METODE PENELITIAN

3.1 Tempat Penelitian

Tempat penelitian dalam studi ini adalah SMP NU Medan, yang menjadi lokasi utama dalam penerapan algoritma K-Means Clustering untuk mengelompokkan data siswa berdasarkan prestasi akademik. Pemilihan SMP NU Medan sebagai tempat penelitian didasarkan pada kebutuhan sekolah dalam mengelola dan menganalisis data siswa secara lebih efektif, khususnya dalam bidang akademik. Dengan memanfaatkan algoritma K-Means, penelitian ini bertujuan untuk membantu pihak sekolah dalam mengidentifikasi pola prestasi siswa sehingga dapat dijadikan dasar dalam perumusan strategi pembelajaran yang lebih tepat sasaran. Selain itu, penelitian ini diharapkan mampu memberikan kontribusi nyata dalam mendukung proses pengambilan keputusan di lingkungan sekolah secara lebih berbasis data.

3.2 Teknik Pengumpulan Data

Dalam penelitian yang berjudul Penerapan Algoritma K-Means Clustering untuk Mengelompokkan Data Siswa SMP NU Medan Berdasarkan Prestasi Akademik, teknik pengumpulan data yang digunakan adalah teknik dokumentasi. Teknik ini dilakukan dengan cara mengumpulkan data nilai akademik siswa dari semester 1 hingga semester 4 yang diperoleh langsung dari pihak administrasi SMP NU Medan. Data yang dikumpulkan meliputi nama siswa, NIS, NISN, serta nilai dari beberapa mata pelajaran inti seperti Matematika, Bahasa Indonesia, Bahasa Inggris, PAI, IPA dll. Penggunaan teknik dokumentasi ini dianggap tepat karena

data yang dibutuhkan bersifat kuantitatif dan telah tersedia dalam bentuk dokumen resmi sekolah. Data tersebut selanjutnya digunakan sebagai bahan utama dalam proses pengolahan dan analisis menggunakan algoritma K-Means untuk mengelompokkan siswa berdasarkan tingkat prestasi akademik mereka.

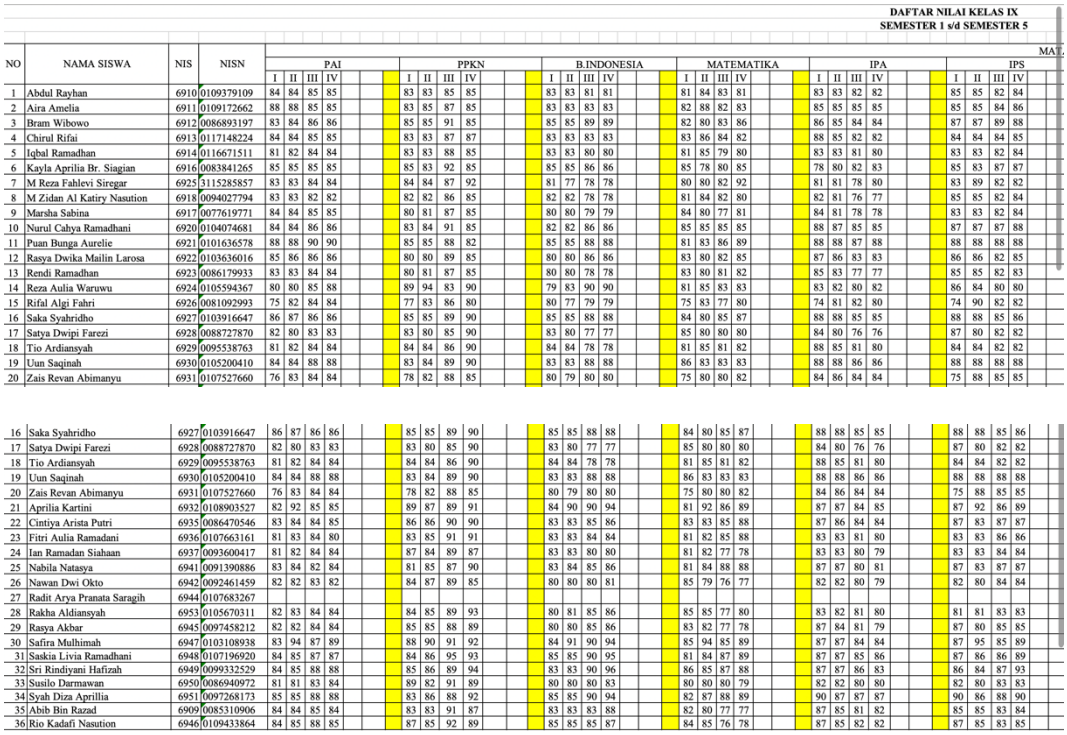
3.3 Data yang Digunakan

Penelitian ini menggunakan data nilai akademik siswa kelas VIII SMP NU Medan yang diperoleh dari laporan hasil belajar semester 1, 2, 3, dan 4. Data tersebut mencakup identitas siswa seperti nomor induk siswa (NIS), nomor induk siswa nasional (NISN), dan nama siswa. Selain itu, data berisi nilai dari berbagai mata pelajaran inti dan muatan lokal, yaitu Pendidikan Agama Islam (PAI), Pendidikan Pancasila dan Kewarganegaraan (PPKN), Bahasa Indonesia, Matematika, Ilmu Pengetahuan Alam (IPA), Ilmu Pengetahuan Sosial (IPS), Bahasa Inggris, Seni Budaya, Pendidikan Jasmani, Olahraga dan Kesehatan (PJOK), Prakarya, IQRA, dan Teknologi Informasi dan Komunikasi (TIK). Setiap mata pelajaran memiliki nilai dari empat semester pertama, yang dijadikan sebagai atribut utama dalam proses pengelompokan siswa.

Penggunaan data nilai dari empat semester bertujuan untuk memberikan gambaran yang lebih stabil dan menyeluruh terhadap prestasi akademik siswa, sehingga proses clustering menggunakan algoritma K-Means dan Fuzzy Tsukamoto dapat menghasilkan pengelompokan yang akurat dan representatif. Data ini dipilih karena mencerminkan perkembangan akademik siswa secara berkelanjutan dalam jangka waktu dua tahun ajaran. Dengan mengolah data ini, penelitian bertujuan untuk mengelompokkan siswa ke dalam beberapa kategori berdasarkan kemiripan nilai akademik, sehingga dapat membantu pihak sekolah

dalam menyusun strategi pembelajaran, bimbingan, dan pemberdayaan potensi siswa secara tepat sasaran.

Tabel 3. 1 Data siswa SMP NU Medan



DAFTAR NILAI KELAS IX
SEMESTER 1 s/d SEMESTER 5

NO	NAMA SISWA	NIS	NISN	PAI				PPKN				B.INDONESIA				MATEMATIKA				IPA				IPS				MAT.
				I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	
1	Abdul Rayhan	6910	0109379109	84	84	85	85	83	83	85	85	83	83	81	81	81	84	83	81	83	83	82	82	85	85	82	84	
2	Aira Amelia	6911	0109172662	88	88	85	85	83	85	87	85	83	83	83	83	82	88	82	83	85	85	85	85	85	85	84	86	
3	Bram Wibowo	6912	0086893197	83	84	86	86	85	85	91	85	85	85	89	89	82	80	83	86	86	85	84	84	87	87	89	88	
4	Chirul Rifai	6913	0117148224	84	84	85	85	83	83	87	87	83	83	83	83	83	86	84	82	88	85	82	82	84	84	84	85	
5	Iqbal Ramadhan	6914	0116671511	81	82	84	84	83	83	88	85	83	83	80	80	81	85	79	80	83	83	81	80	83	83	82	84	
6	Kayla Aprilia Br. Siagian	6916	0083841265	85	85	85	85	85	83	92	85	85	85	86	86	85	78	80	85	78	80	82	83	85	83	87	87	
7	M Reza Fahlevi Siregar	6925	3115285857	83	83	84	84	84	84	87	92	81	77	78	78	80	80	82	92	81	81	78	80	83	89	82	82	
8	M Zidan Al Katiry Nasution	6918	0094027794	83	83	82	82	82	82	86	85	82	82	78	78	81	84	82	80	82	81	76	77	85	85	82	84	
9	Marsha Sabina	6917	0077619771	84	84	85	85	80	81	87	85	80	80	79	79	84	80	77	81	84	81	78	78	83	83	82	84	
10	Nurul Cahya Ramadhani	6920	0104074681	84	84	86	86	83	84	91	85	82	82	86	86	85	85	85	85	88	87	85	85	87	87	87	88	
11	Puan Bunga Aurelie	6921	0101636578	88	88	90	90	85	85	88	82	85	85	88	88	81	83	86	89	88	88	87	88	88	88	88	88	
12	Rasya Dwika Mailin Larosa	6922	0103636016	85	86	86	86	80	80	89	85	80	80	86	86	83	80	82	85	87	86	83	83	86	86	82	85	
13	Rendi Ramadhan	6923	0086179933	83	83	84	84	80	81	87	85	80	80	78	78	83	80	81	82	85	83	77	77	85	85	82	83	
14	Reza Aulia Waruwu	6924	0105594367	80	80	85	88	89	94	83	90	79	83	90	90	81	85	83	83	83	82	80	82	86	84	80	80	
15	Rifal Algi Fahri	6926	0081092993	75	82	84	84	77	83	86	80	80	77	79	79	75	83	77	80	74	81	82	80	74	90	82	82	
16	Saka Syahridho	6927	0103916647	86	87	86	86	85	85	89	90	85	85	88	88	84	80	85	87	88	88	85	85	88	88	85	86	
17	Satya Dwipi Farezi	6928	0088727870	82	80	83	83	83	80	85	90	83	80	77	77	85	80	80	80	84	80	76	76	87	80	82	82	
18	Tio Ardiansyah	6929	0095538763	81	82	84	84	84	84	86	90	84	84	78	78	81	85	81	82	88	85	81	80	84	84	82	82	
19	Uun Saqinah	6930	0105200410	84	84	88	88	83	84	89	90	83	83	88	88	86	83	83	83	88	88	86	86	88	88	88	88	
20	Zais Revan Abimanyu	6931	0107527660	76	83	84	84	78	82	88	85	80	79	80	80	75	80	80	82	84	86	84	84	75	88	85	85	
21	Aprilia Kartini	6932	0108903527	82	92	85	85	89	87	89	91	84	90	90	94	81	92	86	89	87	87	84	85	87	92	86	89	
22	Cintiya Arista Putri	6935	0086470546	83	84	84	85	86	86	90	90	83	83	85	86	83	83	85	88	87	86	84	84	87	83	87	87	
23	Fitri Aulia Ramadani	6936	0107663161	81	83	84	80	83	85	91	91	83	83	84	84	81	82	85	88	83	83	81	80	83	83	86	86	
24	Ian Ramadan Siahaan	6937	0093600417	81	82	84	84	87	84	89	87	83	83	80	80	81	82	77	78	83	83	80	79	83	83	84	84	
25	Nabila Natasya	6941	0091390886	83	84	82	84	81	85	87	90	83	84	85	86	81	84	88	88	87	87	80	81	87	83	87	87	
26	Nawan Dwi Okto	6942	0092461459	82	82	83	82	84	87	89	85	80	80	80	81	85	79	76	77	82	82	80	79	82	80	84	84	
27	Radit Arya Pranata Saragih	6944	0107683267																									
28	Rakha Aldiansyah	6953	0105670311	82	83	84	84	84	85	89	93	80	81	85	86	85	85	77	80	83	82	81	80	81	81	83	83	
29	Rasya Akbar	6945	0097458212	82	82	84	84	85	85	88	89	80	80	85	86	83	82	77	78	87	84	81	79	87	80	85	85	
30	Safira Mulhimah	6947	0103108938	83	94	87	89	88	90	91	92	84	91	90	94	85	94	85	89	87	87	84	84	87	95	85	89	
31	Saskia Livia Ramadhani	6948	0107196920	84	85	87	87	84	86	95	93	85	85	90	95	81	84	87	89	87	87	85	86	87	86	86	89	
32	Sri Rindiyani Hafizah	6949	0099332529	84	85	88	88	85	86	89	94	83	83	90	96	86	85	87	88	87	87	86	83	86	84	87	93	
33	Susilo Darmawan	6950	0086940972	81	81	83	84	89	82	91	89	80	80	80	83	80	80	80	79	82	82	80	80	82	80	83	83	
34	Syah Diza Aprillia	6951	0097268173	85	85	88	88	83	86	88	92	85	85	90	94	82	87	88	89	90	87	87	87	90	86	88	90	
35	Abib Bin Razad	6909	0085310906	84	84	85	84	83	83	91	87	83	83	83	88	82	80	77	77	87	85	81	82	85	85	83	84	
36	Rio Kadafi Nasution	6946	0109433864	84	85	88	85	87	85	92	89	85	85	85	87	84	85	76	78	87	85	82	82	87	85	83	85	

3.4 Tahapan Penelitian

Tahapan penelitian dalam perbandingan algoritma K-Means clustering dan fuzy tsukamoto untuk mengelompokkan data siswa berdasarkan prestasi akademik diawali dengan pengumpulan data nilai siswa SMP NU Medan dari semester 1 hingga semester 4, yang kemudian diproses dalam tahap preprocessing seperti pembersihan data dan penghitungan nilai rata-rata per mata pelajaran. Penelitian ini bertujuan membandingkan kinerja algoritma K-Means Clustering dan Fuzzy Tsukamoto dalam mengelompokkan data siswa berdasarkan prestasi akademik. Alur tahapan penelitian meliputi:

1. Pengumpulan Data

Data nilai siswa SMP NU Medan dikumpulkan mulai dari semester 1 hingga semester 4. Data mencakup nilai tiap mata pelajaran yang akan menjadi variabel input pada proses pengolahan selanjutnya.

2. Preprocessing Data

   Tahap ini meliputi:

   a. Pembersihan data untuk menghapus nilai kosong, duplikat, atau data yang tidak valid.
   b. Penghitungan nilai rata-rata per mata pelajaran bagi setiap siswa.
   c. Normalisasi data untuk menyamakan skala antar atribut sehingga tidak ada variabel yang mendominasi proses pengolahan.

3. Penentuan Kriteria & Parameter

   a. Menentukan jumlah klaster *K* sesuai tujuan pengelompokan, umumnya tiga kategori: Prestasi Tinggi, Prestasi Sedang, dan Prestasi Rendah.
   b. Menetapkan parameter algoritma, seperti nilai awal centroid untuk K-Means atau bentuk fungsi keanggotaan pada Fuzzy Tsukamoto.

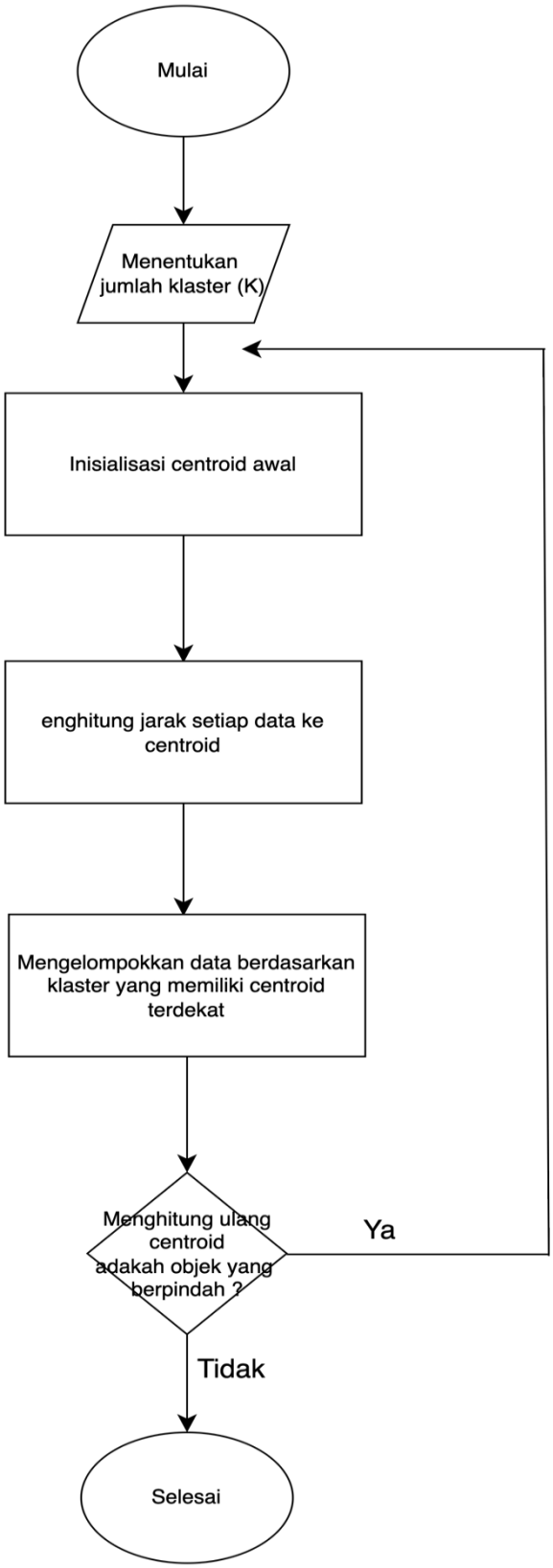
4. Proses dengan Algoritma K-Means

   a. Inisialisasi titik centroid awal.
   b. Hitung jarak setiap data ke semua centroid.
   c. Kelompokkan data ke klaster dengan jarak terdekat.
   d. Perbarui posisi centroid berdasarkan rata-rata anggota klaster.
   e. Ulangi hingga posisi centroid stabil (konvergen).
   f. Mengulangi Proses Hingga Tidak Ada Perubahan Cluster

5. Proses dengan Algoritma Fuzzy Tsukamoto

a. Fuzzifikasi: mencari nilai rata-rata siswa ke dalam himpunan linguistik (Rendah, Sedang, Tinggi).
b. Terapkan basis aturan IF–THEN yang telah ditetapkan.
c. Inferensi: Hitung nilai alpa prdikat (α) untuk setiap aturan dan tentukan keluaran crisp parsial/nilai z (z)
d. Defuzzifikasi menggunakan rata-rata terbobot untuk menghasilkan skor akhir prestasi.
e. Pemetaan skor ke kategori prestasi sesuai ambang batas.

6. Analisis Hasil & Perbandingan
   a. Bandingkan hasil pengelompokan K-Means dan Fuzzy Tsukamoto menggunakan metrik evaluasi seperti Silhouette Coefficient, Cohen's Kappa, atau Adjusted Rand Index.
   b. Interpretasikan perbedaan hasil untuk melihat kelebihan dan kelemahan masing-masing metode dalam konteks data siswa.



Gambar 3. 1 Tahapan K means

Berikut keterangan dari gambar di atas :

1. Menentukan Jumlah Cluster (K)

   Pada tahap awal, peneliti menentukan jumlah cluster (K) yang akan digunakan dalam pengelompokan data siswa SMP NU Medan. Berdasarkan tujuan penelitian untuk mengelompokkan siswa berdasarkan prestasi akademik, maka dipilihlah tiga cluster yaitu siswa dengan prestasi tinggi, sedang, dan rendah. Pemilihan nilai K=3 ini didasarkan pada pertimbangan pedagogis dan pola umum dalam klasifikasi akademik yang digunakan di sekolah, sehingga hasil pengelompokan dapat lebih mudah diinterpretasikan oleh pihak sekolah.

2. Menentukan Titik Pusat Awal (Centroid)

   Setelah jumlah cluster ditentukan, langkah selanjutnya adalah memilih tiga titik pusat awal (centroid) secara acak dari data nilai siswa SMP NU Medan. Nilai-nilai yang digunakan meliputi nilai mata pelajaran inti seperti Matematika, Bahasa Indonesia, Bahasa Inggris, dan IPA dari semester 1 hingga semester 4. Setiap centroid awal merepresentasikan kelompok awal siswa dengan karakteristik nilai tertentu, dan akan menjadi acuan dalam proses pengelompokan awal.

3. Menghitung Jarak Setiap Siswa ke Centroid

   Pada tahap ini, dilakukan perhitungan jarak antara setiap siswa dengan masing-masing centroid menggunakan rumus Euclidean Distance. Misalnya, jika seorang siswa memiliki nilai tinggi dalam semua mata pelajaran, maka jaraknya kemungkinan besar akan lebih dekat ke centroid kelompok berprestasi tinggi. Proses ini dilakukan untuk seluruh data siswa SMP NU

Medan, sehingga setiap siswa dapat diketahui seberapa dekat ia dengan masing-masing cluster yang telah ditentukan.

4. Mengelompokkan Data Siswa ke Cluster Terdekat

   Berdasarkan hasil perhitungan jarak, setiap siswa SMP NU Medan kemudian dikelompokkan ke dalam cluster yang memiliki jarak terdekat dengan nilai centroid. Misalnya, siswa dengan nilai konsisten tinggi akan masuk ke cluster prestasi tinggi, sementara siswa dengan nilai fluktuatif atau rendah akan masuk ke cluster sedang atau rendah. Pengelompokan awal ini memberikan gambaran awal tentang distribusi prestasi akademik di lingkungan sekolah.

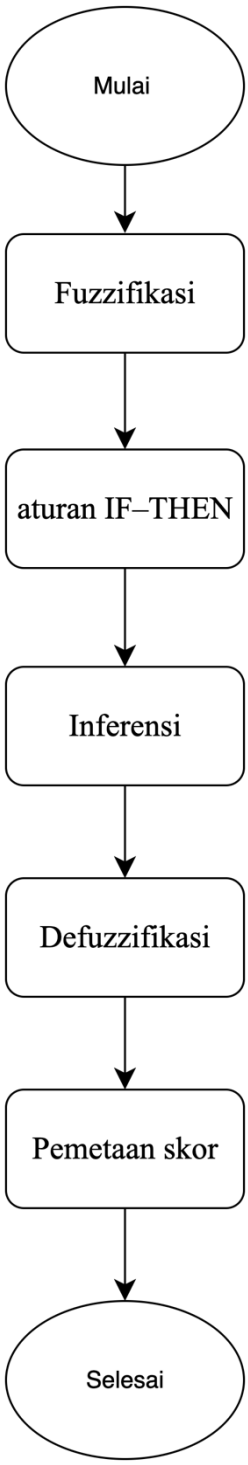
5. Menghitung Ulang Titik Pusat (Centroid) Tiap Cluster

   Setelah pengelompokan awal selesai, dilakukan perhitungan ulang terhadap centroid setiap cluster. Proses ini dilakukan dengan menghitung rata-rata nilai seluruh siswa yang masuk dalam satu cluster. Misalnya, cluster dengan banyak siswa berprestasi akan menghasilkan centroid baru dengan nilai rata-rata tinggi. Proses ini penting untuk menyesuaikan posisi centroid agar lebih representatif terhadap anggota kelompoknya di SMP NU Medan.

6. Mengulangi Proses Hingga Tidak Ada Perubahan Cluster

   Langkah terakhir adalah mengulangi proses perhitungan jarak, pengelompokan, dan pembaruan centroid hingga tidak ada lagi perpindahan siswa antar cluster, atau dengan kata lain, hasil pengelompokan sudah stabil. Dalam konteks data siswa SMP NU Medan, proses ini dilakukan hingga semua siswa tetap berada dalam kelompok prestasi yang sama selama beberapa iterasi berturut-turut. Hasil akhir ini kemudian digunakan sebagai

dasar analisis untuk melihat distribusi siswa berdasarkan kategori prestasi akademik.



Gambar 3. 2 Tahapan Fuzy Tsukamoto

Berikut keterangan dari gambar di atas :

a. Fuzzifikasi

Pada tahap fuzzifikasi, nilai rata-rata setiap siswa yang telah dihitung dari seluruh mata pelajaran diubah ke dalam bentuk linguistik seperti Rendah, Sedang, dan Tinggi berdasarkan fungsi keanggotaan yang telah dirancang. Proses ini memungkinkan data numerik diinterpretasikan secara kualitatif sehingga dapat digunakan dalam sistem inferensi fuzzy. Pemilihan bentuk fungsi keanggotaan misalnya segitiga atau trapezium disesuaikan dengan distribusi data dan kebijakan penilaian sekolah.

b. Penerapan Basis Aturan IF–THEN

Setelah nilai siswa terfuzzifikasi, sistem menerapkan seperangkat aturan IF–THEN yang telah ditentukan sebelumnya bersama pihak ahli atau guru. Setiap aturan menghubungkan kondisi input (misalnya "Nilai Kognitif Tinggi" dan "Nilai Keterampilan Sedang") dengan konsekuen output (*Prestasi Tinggi*). Aturan-aturan ini menjadi kerangka logika yang mengarahkan proses inferensi untuk menghasilkan keluaran sesuai hubungan antar variabel.

c. Inferensi

Tahap inferensi dimulai dengan menghitung nilai α-predikat (α) untuk setiap aturan, yang merepresentasikan tingkat kebenaran kondisi pada bagian premis aturan tersebut. Nilai α diperoleh melalui operasi t-norm seperti *minimum* untuk menggabungkan derajat keanggotaan tiap input. Selanjutnya, setiap α digunakan untuk menentukan keluaran crisp parsial atau nilai zzz dengan membalik fungsi keanggotaan konsekuen yang bersifat monoton.

d. Defuzzifikasi

Defuzzifikasi dilakukan untuk mengubah seluruh keluaran crisp parsial (zzz) yang dihasilkan oleh aturan-aturan menjadi satu nilai akhir yang merepresentasikan skor prestasi siswa. Metode yang digunakan adalah rata-rata terbobot, di mana setiap zzz diberi bobot sesuai dengan nilai α-nya. Hasil perhitungan ini berupa nilai tunggal (crisp) yang menjadi representasi kuantitatif dari prestasi siswa.

e. Pemetaan Skor ke Kategori Prestasi

Tahap terakhir adalah memetakan nilai crisp hasil defuzzifikasi ke dalam kategori prestasi yang telah ditetapkan, seperti *Prestasi Rendah*, *Prestasi Sedang*, atau *Prestasi Tinggi*. Proses pemetaan ini dilakukan berdasarkan ambang batas yang ditentukan sesuai kebijakan sekolah atau analisis data historis. Hasil kategorisasi ini memudahkan pihak sekolah dalam mengevaluasi, mengelompokkan, dan merencanakan strategi pembelajaran yang sesuai dengan tingkat capaian siswa.

3.5 Tools dan Software yang Digunakan

Dalam penelitian ini, peneliti menggunakan beberapa tools dan software untuk mendukung proses pengolahan data dan penerapan algoritma K-Means Clustering. Software utama yang digunakan adalah Python, karena memiliki berbagai library yang powerful untuk pengolahan data dan penerapan algoritma machine learning, seperti Pandas, NumPy. Pandas digunakan untuk mengelola dan memproses data nilai siswa SMP NU Medan dalam bentuk tabel, NumPy digunakan untuk perhitungan matemati. Selain itu, Microsoft Excel juga digunakan sebagai alat bantu awal untuk mengelompokkan, membersihkan, dan menyesuaikan format data yang diperoleh dari pihak sekolah. Seluruh proses dilakukan pada

sistem operasi Windows 10, dengan dukungan aplikasi visual studio code sebagai media utama dalam penulisan dan pelaksanaan kode program.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Deskripsi Data

Data penelitian diperoleh dari SMP NU Medan berupa nilai rapor siswa kelas VIII semester I–IV. Data mencakup 36 siswa dengan identitas NIS, NISN, dan nama siswa, serta nilai 12 mata pelajaran inti yaitu:

1. Pendidikan Agama Islam (PAI)
2. Pendidikan Pancasila dan Kewarganegaraan (PPKN)
3. Bahasa Indonesia
4. Matematika
5. Ilmu Pengetahuan Alam (IPA)
6. Ilmu Pengetahuan Sosial (IPS)
7. Bahasa Inggris
8. Seni Budaya
9. Pendidikan Jasmani, Olahraga, dan Kesehatan (PJOK)
10. Prakarya
11. IQRA
12. TIK

Setiap mata pelajaran memiliki nilai pada empat semester (I–IV). Sebelum digunakan, data melalui tahap preprocessing berupa: Pembersihan data, Perhitungan Rata-rata dan Normalisasi data.

4.1.1 Karakteristik Dataset

Dataset awal memiliki karakteristik sebagai berikut:

a. Jumlah siswa: 36 siswa
b. Jumlah mata pelajaran: 12 mata pelajaran
c. Periode penilaian: 4 semester (I, II, III, IV)
d. Total data nilai: 1.728 data point (36 siswa × 12 mata pelajaran × 4 semester)

4.1.2 Preprocessing Data

Sebelum dilakukan pengolahan dengan algoritma clustering, data melewati beberapa tahap preprocessing:

1. Pembersihan Data
    a. Identifikasi dan penanganan missing values atau data yang kosong
    b. Dari analisis data, ditemukan beberapa nilai kosong pada siswa tertentu (contoh: Radit Arya Pranata Saragih dengan NIS 6944 tidak memiliki nilai sama sekali)
    c. Siswa dengan data tidak lengkap akan diexclude atau tidak dimaksukan dari analisis

4.2 Perhitungan Manual Algoritma K-Means

1. Ambil data fitur rata-rata nilai per siswa dari mata pelajaran: PAI, PPKN, B.Indonesia, Matematika, IPA, IPS, B.Inggris, Seni Budaya, PJOK, Prakarya, Iqra, TIK.

   Contoh siswa Abdul Rayhan

Mata Pelajaran	Nilai Semester I -IV	Total Nilai

PAI	(84+84+85+85)/4	84.5
PPKN	(83+83+85+85)/4	84.0
B. Indonesia	(83+83+81+81)/4	82.0
Matematika	(81+84+83+81)/4	82.25
IPA	(83+83+82+82)/4	82.5
IPS	(85+85+82+84)/4	84.0
B. Inggris	(83+80+85+83)/4	82.75
Seni Budaya	(84+84+88+85)/4	85.25
PJOK	(88+84+88+85)/4	86.25
Prakarya	(86+83+88+84)/4	85.25
IQRA	(83+84+85+85)/4	84.25
TIK	(82+83+88+94)/4	86.75
Rata-rata Nilai		**84.1**

2. Tentukan jumlah cluster : Tinggi, Sedang, Rendah
3. Inisialisasi centroid awal
   a. C1 (Prestasi Rendah) : 75
   b. C2 (Prestasi Sedang) : 85
   c. C3 (Prestasi Tinggi) : 95
4. Hitung jarak Euclidean tiap siswa ke centroid

   Contoh Abdul Rayhan = 84.1

   Ket : d(c) = jarak centroid

   a. d(C1) = 84.1 – 75 = 9.0
   b. d(C2) = 84.1 – 85 = -0.9

c. d(C3) = 84.1 – 95 = -10.9

Masuk cluster C2 (Sedang) karena 0.9

5. Jadi Abdul rayhan masuk cluster C2 (Sedang) karena jarak paling kecil
6. Update centroid
   a. Hasil assignment (langkah 4–5): Abdul Rayhan masuk C2 (Sedang) karena 84,1 − 85 = **0,9** paling kecil.
   b. Anggota cluster:
      - C1 = 75
      - C2 = 84.1
      - C3 = 95
7. Iterasi sampai centroid stabil (tidak berubah lagi).
   a. Ulangi assignment dengan centroid hasil update:
      - d(C1) = 84.1 – 75 = 9.0
      - d(C2) = 84.1 – 84.1 = 0.0
      - d(C3) = 84.1 – 95 = 10.9

Tetap masuk C2.

C2 tetap rata-rata = 84,1; C1 dan C3 tetap (tidak ada anggota).

Karena keanggotaan dan centroid tidak berubah, algoritma konvergen/stabil di sini.

1. Contoh Siswa Aira Amelia

Mata Pelajaran	Nilai Semester I -IV	Total Nilai
PAI	88+88+85+85/4	86,50
PPKN	83+85+87+85/4	85,00
B. Indonesia	83+83+83+83/4	83,00
Matematika	82+88+82+83/4	83,75

IPA	85+85+85+85/4	85,00
IPS	85+85+84+86/4	85,00
B. Inggris	85+83+86+894	85,75
Seni Budaya	88+87+95+85/4	88,75
PJOK	88+87+90+87/4	88,00
Prakarya	90+82+95+85/4	88,00
IQRA	84+84+87+84/4	84,75
TIK	83+85+88+93/4	87,25
Rata-rata Nilai		**85.90**

2. Tentukan jumlah cluster : Tinggi, Sedang, Rendah
3. Inisialisasi centroid awal
   a. C1 (Prestasi Rendah) : 75
   b. C2 (Prestasi Sedang) : 85
   c. C3 (Prestasi Tinggi) : 95
4. Hitung jarak Euclidean tiap siswa ke centroid

   Contoh Aira Amelia = 85.90

   Ket : d(c) = jarak centroid

   a. d(C1) = 85.90 – 75 = 10.90
   b. d(C2) = 85.90 – 85 = 0.9
   c. d(C3) = 85.90 – 95 = 9.10

   Jarak terkecil = 0.90 maka masuk ke cluster C2 (Sedang)
5. Jadi Aira Amelia masuk cluster C2 (Sedang) karena jarak paling kecil
6. Update centroid

Keterangan :

- jika hanya Aira Amelia yang masuk C2 maka centroid c2 tetap 85,90
- jika Aira Amelia dan Abdul Rayhan masuk C2 maka :

C2 baru = $\frac{84.10+85.90}{2}$ = 85.00

Anggota cluster:

- C1 = 75
- C2 = 85
- C3 = 95

7. Iterasi sampai centroid stabil (tidak berubah lagi).

Ulangi assignment dengan centroid hasil update:

- d(C1) = 85.90 – 75 = 10.9
- d(C2) = 85.90 – 85 = 0.9
- d(C3) = 85.90 – 95 = 9.1

Tetap masuk C2. Update centroid lagi:

Karena keanggotaan dan centroid tidak berubah, algoritma konvergen/stabil di sini.

1. Contoh siswa Bram Wibowo

Mata Pelajaran	Nilai Semester I -IV	Total Nilai
PAI	83+84+86+86/4	84,75
PPKN	85+85+91+85/4	86,50
B. Indonesia	85+85+89+89/4	87,00
Matematika	82+80+83+86/4	82,75
IPA	86+85+84+84/4	84,75
IPS	87+87+89+88/4	87,75

B. Inggris	85+87+88+86/4	86,50
Seni Budaya	83+84+87+85/4	84,75
PJOK	82+84+93+84/4	85,75
Prakarya	84+84+87+87/4	85,50
IQRA	83+85+86+86/4	85,00
TIK	82+83+88+94/4	86,75
Rata-rata Nilai		**85.65**

2. Tentukan jumlah cluster : Tinggi, Sedang, Rendah
3. Inisialisasi centroid awal
   a. C1 (Prestasi Rendah) : 75
   b. C2 (Prestasi Sedang) : 85
   c. C3 (Prestasi Tinggi) : 95
4. Hitung jarak Euclidean tiap siswa ke centroid

   Contoh Bram Wibowo = 85.65

   Ket : d(c) = jarak centroid

   a. d(C1) = 85.65 – 75 = 10.65
   b. d(C2) = 85.65 – 85 = 0.65
   c. d(C3) = 85.65 – 95 = 9.35

   Jarak terkecil = 0.65 maka masuk ke cluster C2 (Sedang)
5. Jadi Bram Wibowo masuk cluster C2 (Sedang) karena jarak paling kecil
6. Update centroid

   Keterangan :

   - jika hanya Bram Wibowo yang masuk C2 maka centroid c2 tetap 85,65

- jika Aira Amelia dan Abdul Rayhan masuk C2 maka :

C2 baru = $\frac{84.10+85.90+85.65}{2}$ = 85.22

Anggota cluster:

- C1 = 75
- C2 = 85.22
- C3 = 95

7. Iterasi sampai centroid stabil (tidak berubah lagi).

Ulangi assignment dengan centroid hasil update:

- d(C1) = 85.65 – 75 = 10.65
- d(C2) = 85.65 – 85.22 = 0.43
- d(C3) = 85.65 – 95 = 9.35

Tetap masuk C2.

Karena keanggotaan dan centroid tidak berubah, algoritma konvergen/stabil di sini. Lakukan perhitungan yang sama untuk seluruh siswa guna memperoleh cluster masing-masing siswa.

4.3 Perhitungan Manual Algoritma Fuzzy Tsukamoto

1. Ambil nilai per mata pelajaran chirul Rifal

Mata Pelajaran	Nilai Semester I -IV
PAI	84, 84, 85, 85
PPKN	83, 83, 87, 87
B. Indonesia	83, 83, 83, 83

Matematika	83, 86, 84, 82
IPA	88, 85, 82, 82
IPS	84, 84, 84, 85
B. Inggris	83, 85, 86, 86
Seni Budaya	81, 82, 87, 85
PJOK	82, 84, 89, 87
Prakarya	84, 84, 95, 84
IQRA	84, 85, 85, 85
TIK	84, 84, 88, 90

2. Hitung rata-rata tiap mapel

Rumus: rata-rata mata pelajaran $= \frac{nilai\ 1+nilai\ 2+nilai3+nilai4}{4}$

Contoh:

- PAI = (84+84+85+85)/4 = 84,5
- PPKN = (83+83+87+87)/4 = 85,0
- B. Indonesia = (83+83+83+83)/4 = 83,0
- Matematika = (83+86+84+82)/4 = 83,75
- IPA = (88+85+82+82)/4 = 84,25
- IPS = (84+84+84+85)/4 = 84,25
- B. Inggris = (83+85+86+86)/4 =85,0
- Seni Budaya = (81+82+87+85)/4 = 83,75
- PJOK = (82+84+89+87)/4 = 85,5
- Prakarya = (84+84+95+84)/4 = 86,75

- IQRA = (84+85+85+85)/4 = 84,75
- TIK = (84+84+88+90)/4 = 86,5

ambil rata-rata mapel yang sudah dihitung, lalu rata-ratakan lagi:

$$= \frac{84.5 + 85.0 + 83.0 + 83.75 + 84.25 + 84.25 + 85.0 + 83.75 + 85.5 + 86.75 + 84.75 + 86.5}{12}$$

$$= 85{,}58$$

3. Ubah ke bobot (aturan fuzzy sederhana Anda)

$$\text{Bobot} = \begin{cases} 0 & x < 60\ (Rendah) \\ 5 & 60 \leq x < 79\ (Sedang) \\ 10 & x \geq 80\ (Tinggi) \end{cases}$$

Kalau nilai rata-rata 0 - 60 maka bobot = 0

Kalau 60–79 maka bobot = 5

Kalau 80–100 maka bobot = 10

4. Karena **85.8 ≥ 80**, maka kategori = **Tinggi (bobot 10)**.

   Bobot fuzzy = **10**.

5. Interpretasi

Hasil Fuzzy Tsukamoto untuk chirul Rifal:

- Rata-rata Nilai: 84.8
- Kategori**:** Tinggi
- Bobot**:** 10

Untuk mendapatkan bobot fuzzy (0, 5, 10) setiap siswa, maka perhitungan yang sama harus dilakukan untuk seluruh siswa. Jadi, setiap siswa dihitung rata-rata per mata pelajaran (dari Semester I–IV), lalu dari 12 mata pelajaran dirata-ratakan lagi untuk memperoleh rata-rata keseluruhan. Rata-rata inilah yang digunakan untuk menentukan kategori (Rendah, Sedang, Tinggi) sesuai aturan fuzzy yang sudah kita tetapkan.

4.4 Hasil Implementasi Algoritma K-Means dengan Program

Implementasi algoritma K-Means pada data nilai siswa kelas IX menghasilkan pengelompokan ke dalam tiga kategori, yaitu Rendah, Sedang, dan Tinggi, sesuai dengan jarak terdekat terhadap centroid awal (75, 85, dan 95). Proses perhitungan dilakukan dengan menghitung rata-rata setiap mata pelajaran dari semester I sampai IV, kemudian dirata-ratakan lagi menjadi satu nilai akhir. Nilai akhir inilah yang digunakan sebagai dasar penentuan cluster. Dari hasil perhitungan, terlihat bahwa sebagian besar siswa masuk ke dalam kategori Sedang, karena rata-rata nilai mereka berada di sekitar centroid 85.

Hasil pengelompokan menunjukkan bahwa kategori Tinggi ditempati oleh siswa dengan rata-rata nilai di atas 87, seperti Syah Diza Aprillia, Sri Rindiyani Hafizah, dan Aprilia Kartini. Sementara itu, kategori Rendah diisi oleh siswa dengan rata-rata nilai di bawah 80, seperti Zais Revan Abimanyu, Nurul Cahya Ramadhani, dan Safira Mulhimah. Adapun siswa yang datanya tidak lengkap,

seperti Radit Arya Pranata Saragih, secara otomatis dikategorikan sebagai “Data tidak lengkap” dan tidak dapat dipetakan dalam cluster. Dengan demikian, hasil implementasi program K-Means ini mampu memberikan gambaran yang lebih objektif mengenai distribusi prestasi siswa berdasarkan data nilai rata-rata.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	NAMA SISWA	PAI_AVG	PPKN_AVG	DONESIA_A	TEMATIKA_	IPA_AVG	IPS_AVG	INGGRIS_AV	BUDAYA_/	PJOK_AVG	AKARYA_A	IQRA_AVG	TIK_AVG	RATA_RATA	KATEGORI	
2	Abdul Rayhan	84,5	84	82	82,25	82,5	84	82,75	85,25	86,25	85,25	84,25	86,75	84,14583333	Sedang	
3	Aira Amelia	86,5	85	83	83,75	85	85	85,75	88,75	88	88	84,75	87,25	85,89583333	Sedang	
4	Bram Wibowo	84,75	86,5	87	82,75	84,75	87,75	86,5	84,75	85,75	85,5	85	86,75	85,64583333	Sedang	
5	Chirul Rifai	84,5	85	83	83,75	84,25	84,25	85	83,75	85,5	86,75	84,75	86,5	84,75	Sedang	
6	Iqbal Ramadhan	82,75	84,75	81,5	81,25	81,75	83	82,5	83,75	84,75	84,5	83,25	85,5	83,27083333	Sedang	
7	Kayla Aprilia Br. Siagian	85	86,25	85,5	82	80,75	85,5	87,25	86,75	85,75	87	84,5	89,5	85,47916667	Sedang	
8	M Reza Fahlevi Siregar	83,5	86,75	78,5	83,5	80	84	81,75	86,25	84,25	84	82,75	87,25	83,54166667	Sedang	
9	M Zidan Al Katiry Nasution	82,5	83,75	80	81,75	79	84	80,75	84,5	83,5	85,25	83	85,5	82,79166667	Sedang	
10	Marsha Sabina	84,5	83,25	79,5	80,5	80,25	83	83,5	85	85,25	85	83,75	86,5	83,33333333	Sedang	
11	Nurul Cahya Ramadhani	85	77,75	76,25	78	78	77,5	80,5	77,5	76,25	81,25	81	80,75	79,14583333	Rendah	
12	Puan Bunga Aurelie	89	85	86,5	84,75	87,75	88	89,25	88,5	88,5	88	88	87,25	87,54166667	Sedang	
13	Rasya Dwika Mailin Larosa	85,75	83,5	83	82,5	84,75	84,75	84,5	85,5	85,25	85,25	84,5	87,75	84,75	Sedang	
14	Rendi Ramadhan	83,5	83,25	79	81,5	80,5	83,75	82	84,75	84,5	85	83,75	86	83,125	Sedang	
15	Reza Aulia Waruwu	83,25	89	85,5	83	81,75	82,5	81,25	83,5	81,25	82,75	22,25	62,75	76,5625	Rendah	
16	Rifal Algi Fahri	81,25	81,5	78,75	78,75	79,25	82	80,5	78,25	81,5	81,25	62,5	65	77,54166667	Rendah	
17	Saka Syahridho	86,25	87,25	86,5	84	86,5	86,75	86,5	89	86,75	87,5	85,5	87,5	86,66666667	Sedang	
18	Satya Dwipi Farezi	82	84,5	79,25	81,25	79	82,75	82,25	83	84	84,25	82,5	85,75	82,54166667	Sedang	
19	Tio Ardiansyah	77,25	86	75	79	79,25	72	74	74	76,75	76	75,5	73,25	76,5	Rendah	
20	Uun Saqinah	86	86,5	85,5	83,75	87	88	87,25	88,5	87,5	88,75	88	87,25	87	Sedang	
21	Zais Revan Abimanyu	81,75	83,25	79,75	79,25	84,5	83,25	82,75	78,25	85,5	85	62,5	66,25	79,33333333	Rendah	
22	Aprilia Kartini	86	89	89,5	87	85,75	88,5	86,5	91,75	89,5	88,5	83,5	89,5	87,91666667	Sedang	
23	Cintiya Arista Putri	84	88	84,25	84,75	85,25	86	86	87	88	84,25	84,25	86,5	85,6875	Sedang	
24	Fitri Aulia Ramadani	82	82,25	75	79	76,75	75,5	74	71,25	76,75	78,25	74	79	76,97916667	Rendah	
25	Ian Ramadan Siahaan	82,75	86,75	81,5	79,5	81,25	83,5	83,75	85,75	86,75	87	82,75	85,75	83,91666667	Sedang	
26	Nabila Natasya	83,25	85,75	84,5	85,25	83,75	86	84,75	87,25	87,75	87,5	82,5	85,75	85,33333333	Sedang	
27	Nawan Dwi Okto	82,25	86,25	80,25	79,25	80,75	82,5	83	85,75	84,5	86,75	82,25	82,25	82,97916667	Sedang	
28	Radit Arya Pranata Saragih														Data tidak lengkap	
29	Rakha Aldiansyah	83,25	87,75	83	81,75	81,5	82	84	87,25	89,5	86,75	84,5	86,25	84,79166667	Sedang	
30	Rasya Akbar	83	86,75	82,75	80	82,75	84,25	85	87,25	88	88	82	88,25	84,83333333	Sedang	
31	Safira Mulhimah	88,25	81,75	77,75	78,75	78,75	74,75	77,25	76,5	77,5	76,5	77	77,5	78,52083333	Rendah	
32	Saskia Livia Ramadhani	85,75	89,5	88,75	85,25	86,25	87	87,75	92	89,75	90,25	85,75	87	87,91666667	Sedang	
33	Sri Rindiyani Hafizah	86,25	88,5	88	86,5	85,75	87,5	88,25	92	89,25	92	85	87	88	Sedang	
34	Susilo Darmawan	82,25	87,75	80,75	79,75	81	82	82,75	85,25	87,25	87,25	83	86	83,75	Sedang	
35	Syah Diza Aprillia	86,5	87,25	88,5	86,5	87,75	88,5	88,75	92,5	89,25	89,5	85,25	89	88,27083333	Sedang	
36	Abib Bin Razad	84,25	82,5	76,25	79,5	80	74,5	76	77,75	77,5	77,25	79	75,5	78,33333333	Rendah	
37	Rio Kadafi Nasution	85,5	88,25	85,5	80,75	84	85	85	86	87,5	87,25	85	86	85,47916667	Sedang	

	A	B	C
1	NAMA SISWA	RATA_RATA	KATEGORI
2	Syah Diza Aprillia	88,27083333	Sedang
3	Sri Rindiyani Hafizah	88	Sedang
4	Aprilia Kartini	87,91666667	Sedang
5	Saskia Livia Ramadha	87,91666667	Sedang
6	Puan Bunga Aurelie	87,54166667	Sedang
7	Uun Saqinah	87	Sedang
8	Saka Syahridho	86,66666667	Sedang
9	Aira Amelia	85,89583333	Sedang
10	Cintiya Arista Putri	85,6875	Sedang
11	Bram Wibowo	85,64583333	Sedang
12	Rio Kadafi Nasution	85,47916667	Sedang
13	Kayla Aprilia Br. Siagi	85,47916667	Sedang
14	Nabila Natasya	85,33333333	Sedang
15	Rasya Akbar	84,83333333	Sedang
16	Rakha Aldiansyah	84,79166667	Sedang
17	Rasya Dwika Mailin La	84,75	Sedang
18	Chirul Rifai	84,75	Sedang
19	Abdul Rayhan	84,14583333	Sedang
20	Ian Ramadan Siahaan	83,91666667	Sedang
21	Susilo Darmawan	83,75	Sedang
22	M Reza Fahlevi Sirega	83,54166667	Sedang
23	Marsha Sabina	83,33333333	Sedang
24	Iqbal Ramadhan	83,27083333	Sedang
25	Rendi Ramadhan	83,125	Sedang
26	Nawan Dwi Okto	82,97916667	Sedang
27	M Zidan Al Katiry Nas	82,79166667	Sedang
28	Satya Dwipi Farezi	82,54166667	Sedang
29	Zais Revan Abimanyu	79,33333333	Rendah
30	Nurul Cahya Ramadh	79,14583333	Rendah
31	Safira Mulhimah	78,52083333	Rendah
32	Abib Bin Razad	78,33333333	Rendah
33	Rifal Algi Fahri	77,54166667	Rendah
34	Fitri Aulia Ramadani	76,97916667	Rendah
35	Reza Aulia Waruwu	76,5625	Rendah
36	Tio Ardiansyah	76,5	Rendah
37	Radit Arya Pranata Saragih		Data tidak lengkap

Gambar 4. 1 Hasil Pengelompokkan Algoritma Kmeans

4.5 Hasil Implementasi Algoritma Fuzzy Tsukamoto dengan Progam

Implementasi algoritma Fuzzy Tsukamoto dilakukan dengan mengambil rata-rata nilai setiap mata pelajaran dari Semester I hingga IV, kemudian dihitung rata-rata keseluruhan untuk menentukan kategori prestasi siswa. Berdasarkan aturan fuzzy yang telah ditetapkan, nilai rata-rata di bawah 60 dikategorikan Rendah (bobot 0), nilai 60–79 dikategorikan Sedang (bobot 5), dan nilai ≥80 dikategorikan Tinggi (bobot 10). Hasil pengolahan data menggunakan program menunjukkan bahwa sebagian besar siswa masuk dalam kategori Tinggi dengan bobot 10, sedangkan beberapa siswa berada pada kategori Sedang dengan bobot 5, dan hanya sedikit yang terklasifikasi sebagai Rendah.

Program yang diimplementasikan berhasil memberikan pembagian kategori yang lebih tegas dibandingkan K-Means karena aturan fuzzy langsung memetakan nilai rata-rata ke dalam bobot tertentu. Dari hasil keluaran, terlihat siswa dengan rata-rata di atas 80 secara konsisten dikategorikan sebagai **Tinggi**, sedangkan siswa dengan rata-rata antara 60–79 masuk dalam kategori **Sedang**. Dengan pendekatan ini, Fuzzy Tsukamoto dapat memberikan hasil klasifikasi yang mudah dipahami dan sesuai dengan batasan logika fuzzy yang telah ditentukan, sehingga memudahkan pihak sekolah dalam mengelompokkan tingkat prestasi siswa.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	NAMA SISWA	PAI_AVG	PPKN_AVG	IDONESIA_A	EMATIKA_	IPA_AVG	IPS_AVG	INGGRIS_A	BUDAYA_A	PJOK_AVG	AKARYA_A	IQRA_AVG	TIK_AVG	RATA_RATA	KATEGORI	BOBOT	
2	Abdul Rayhan	84,5	84	82	82,25	82,5	84	82,75	85,25	86,25	85,25	84,25	86,75	84,14583333	Tinggi	10	
3	Aira Amelia	86,5	85	83	83,75	85	85	85,75	88,75	88	88	84,75	87,25	85,89583333	Tinggi	10	
4	Bram Wibowo	84,75	86,5	87	82,75	84,75	87,75	86,5	84,75	85,75	85,5	85	86,75	85,64583333	Tinggi	10	
5	Chirul Rifai	84,5	85	83	83,75	84,25	84,25	85	83,75	85,5	86,75	84,75	86,5	84,75	Tinggi	10	
6	Iqbal Ramadhan	82,75	84,75	81,5	81,25	81,75	83	82,5	83,75	84,75	84,5	83,25	85,5	83,27083333	Tinggi	10	
7	Kayla Aprilia Br. Siagian	85	86,25	85,5	82	80,75	85,5	87,25	86,75	85,75	87	84,5	89,5	85,47916667	Tinggi	10	
8	M Reza Fahlevi Siregar	83,5	86,75	78,5	83,5	80	84	81,75	86,25	84,25	84	82,75	87,25	83,54166667	Tinggi	10	
9	M Zidan Al Katiry Nasution	82,5	83,75	80	81,75	79	84	80,75	84,5	83,5	85,25	83	85,5	82,79166667	Tinggi	10	
10	Marsha Sabina	84,5	83,25	79,5	80,5	80,25	83	83,5	85	85,25	85	83,75	86,5	83,33333333	Tinggi	10	
11	Nurul Cahya Ramadhani	85	77,75	76,25	78	78	77,5	80,5	77,5	76,25	81,25	81	80,75	79,14583333	Tinggi	10	
12	Puan Bunga Aurelie	89	85	86,5	84,75	87,75	88	89,25	88,5	88,5	88	88	87,25	87,54166667	Tinggi	10	
13	Rasya Dwika Mailin Larosa	85,75	83,5	83	82,5	84,75	84,75	84,5	85,5	85,25	85,25	84,5	87,75	84,75	Tinggi	10	
14	Rendi Ramadhan	83,5	83,25	79	81,5	80,5	83,75	82	84,75	84,5	85	83,75	86	83,125	Tinggi	10	
15	Reza Aulia Waruwu	83,25	89	85,5	83	81,75	82,5	81,25	83,5	81,25	82,75	22,25	62,75	76,5625	Sedang	5	
16	Rifal Algi Fahri	81,25	81,5	78,75	78,75	79,25	82	80,5	78,25	81,5	81,25	62,5	65	77,54166667	Sedang	5	
17	Saka Syahridho	86,25	87,25	86,5	84	86,5	86,75	86,5	89	86,75	87,5	85,5	87,5	86,66666667	Tinggi	10	
18	Satya Dwipi Farezi	82	84,5	79,25	81,25	79	82,75	82,25	83	84	84,25	82,5	85,75	82,54166667	Tinggi	10	
19	Tio Ardiansyah	77,25	86	75	79	79,25	72	74	74	76,75	76	75,5	73,25	76,5	Sedang	5	
20	Uun Saqinah	86	86,5	85,5	83,75	87	88	87,25	88,5	87,5	88,75	88	87,25	87	Tinggi	10	
21	Zais Revan Abimanyu	81,75	83,25	79,75	79,25	84,5	83,25	82,75	78,25	85,5	85	62,5	66,25	79,33333333	Tinggi	10	
22	Aprilia Kartini	86	89	89,5	87	85,75	88,5	86,5	91,75	89,5	88,5	83,5	89,5	87,91666667	Tinggi	10	
23	Cintiya Arista Putri	84	88	84,25	84,75	85,25	86	86	87	88	84,25	84,25	86,5	85,6875	Tinggi	10	
24	Fitri Aulia Ramadani	82	82,25	75	79	76,75	75,5	74	71,25	76,75	78,25	74	79	76,97916667	Sedang	5	
25	Ian Ramadan Siahaan	82,75	86,75	81,5	79,5	81,25	83,5	83,75	85,75	86,75	87	82,75	85,75	83,91666667	Tinggi	10	
26	Nabila Natasya	83,25	85,75	84,5	85,25	83,75	86	84,75	87,25	87,75	87,5	82,5	85,75	85,33333333	Tinggi	10	
27	Nawan Dwi Okto	82,25	86,25	80,25	79,25	80,75	82,5	83	85,75	84,5	86,75	82,25	82,25	82,97916667	Tinggi	10	
28	Radit Arya Pranata Saragih														Data tidak lengkap		
29	Rakha Aldiansyah	83,25	87,75	83	81,75	81,5	82	84	87,25	89,5	86,75	84,5	86,25	84,79166667	Tinggi	10	
30	Rasya Akbar	83	86,75	82,75	80	82,75	84,25	85	87,25	88	88	82	88,25	84,83333333	Tinggi	10	
31	Safira Mulhimah	88,25	81,75	77,75	78,75	78,75	74,75	77,25	76,5	77,5	76,5	77	77,5	78,52083333	Sedang	5	
32	Saskia Livia Ramadhani	85,75	89,5	88,75	85,25	86,25	87	87,75	92	89,75	90,25	85,75	87	87,91666667	Tinggi	10	
33	Sri Rindiyani Hafizah	86,25	88,5	88	86,5	85,75	87,5	88,25	92	89,25	92	85	87	88	Tinggi	10	
34	Susilo Darmawan	82,25	87,75	80,75	79,75	81	82	82,75	85,25	87,25	87,25	83	86	83,75	Tinggi	10	
35	Syah Diza Aprillia	86,5	87,25	88,5	86,5	87,75	88,5	88,75	92,5	89,25	89,5	85,25	89	88,27083333	Tinggi	10	
36	Abib Bin Razad	84,25	82,5	76,25	79,5	80	74,5	76	77,75	77,5	77,25	79	75,5	78,33333333	Sedang	5	
37	Rio Kadafi Nasution	85,5	88,25	85,5	80,75	84	85	85	86	87,5	87,25	85	86	85,47916667	Tinggi	10	

	A	B	C	D	E
1	NAMA SISWA	RATA_RATA	KATEGORI	BOBOT	
2	Syah Diza Aprillia	88,27083333	Tinggi	10	
3	Sri Rindiyani Hafizah	88	Tinggi	10	
4	Aprilia Kartini	87,91666667	Tinggi	10	
5	Saskia Livia Ramadhani	87,91666667	Tinggi	10	
6	Puan Bunga Aurelie	87,54166667	Tinggi	10	
7	Uun Saqinah	87	Tinggi	10	
8	Saka Syahridho	86,66666667	Tinggi	10	
9	Aira Amelia	85,89583333	Tinggi	10	
10	Cintiya Arista Putri	85,6875	Tinggi	10	
11	Bram Wibowo	85,64583333	Tinggi	10	
12	Rio Kadafi Nasution	85,47916667	Tinggi	10	
13	Kayla Aprilia Br. Siagian	85,47916667	Tinggi	10	
14	Nabila Natasya	85,33333333	Tinggi	10	
15	Rasya Akbar	84,83333333	Tinggi	10	
16	Rakha Aldiansyah	84,79166667	Tinggi	10	
17	Rasya Dwika Mailin Larosa	84,75	Tinggi	10	
18	Chirul Rifai	84,75	Tinggi	10	
19	Abdul Rayhan	84,14583333	Tinggi	10	
20	Ian Ramadan Siahaan	83,91666667	Tinggi	10	
21	Susilo Darmawan	83,75	Tinggi	10	
22	M Reza Fahlevi Siregar	83,54166667	Tinggi	10	
23	Marsha Sabina	83,33333333	Tinggi	10	
24	Iqbal Ramadhan	83,27083333	Tinggi	10	
25	Rendi Ramadhan	83,125	Tinggi	10	
26	Nawan Dwi Okto	82,97916667	Tinggi	10	
27	M Zidan Al Katiry Nasution	82,79166667	Tinggi	10	
28	Satya Dwipi Farezi	82,54166667	Tinggi	10	
29	Zais Revan Abimanyu	79,33333333	Tinggi	10	
30	Nurul Cahya Ramadhani	79,14583333	Tinggi	10	
31	Safira Mulhimah	78,52083333	Sedang	5	
32	Abib Bin Razad	78,33333333	Sedang	5	
33	Rifal Algi Fahri	77,54166667	Sedang	5	
34	Fitri Aulia Ramadani	76,97916667	Sedang	5	
35	Reza Aulia Waruwu	76,5625	Sedang	5	
36	Tio Ardiansyah	76,5	Sedang	5	
37	Radit Arya Pranata Saragih		Data tidak lengkap		

Gambar 4. 2 Hasil Pengelompokkan Algoritma Fuzzy Tsukamoto

4.6 Pembahasan Perbandingan Hasil K-Means dan Fuzzy Tsukamoto

Perbandingan hasil antara algoritma K-Means dan Fuzzy Tsukamoto menunjukkan adanya perbedaan mendasar baik dari sisi konsep maupun keluaran yang dihasilkan. Algoritma K-Means bekerja dengan prinsip *clustering unsupervised* berbasis jarak Euclidean, sehingga posisi centroid akan menyesuaikan dengan distribusi data siswa. Dengan parameter awal centroid 75, 85, dan 95, hasil clustering dapat membagi siswa ke dalam tiga kelompok (Rendah, Sedang, Tinggi) secara lebih proporsional. Sebaliknya, algoritma Fuzzy Tsukamoto menggunakan aturan linguistik sederhana berbasis batas nilai, yaitu <60 sebagai Rendah (bobot 0), 60–79 sebagai Sedang (bobot 5), dan ≥80 sebagai Tinggi (bobot 10). Dengan aturan tersebut, kategorisasi menjadi lebih kaku karena tidak dipengaruhi oleh distribusi data, melainkan sepenuhnya oleh batas nilai yang ditetapkan.

Dari sisi distribusi hasil, K-Means lebih seimbang karena siswa dengan rata-rata sekitar 84–86 dapat masuk kategori Sedang apabila lebih dekat dengan centroid 85. Sementara itu, Fuzzy Tsukamoto langsung mengkategorikan semua siswa dengan rata-rata ≥80 ke dalam kategori Tinggi. Hal ini terlihat pada contoh tiga siswa, yaitu Abdul Rayhan (84,1), Aira Amelia (85,9), dan Chirul Rifai (84,8), yang oleh K-Means dimasukkan ke cluster Sedang karena posisinya dekat dengan centroid 85, sedangkan oleh Fuzzy Tsukamoto dikategorikan sebagai Tinggi karena nilainya melewati batas ≥80. Dengan demikian, K-Means lebih fleksibel dalam mengelompokkan data berdasarkan distribusi, sedangkan Fuzzy Tsukamoto memberikan hasil klasifikasi yang lebih sederhana dan tegas sesuai aturan yang ditentukan.

Perbandingan Konsep

Aspek	K-Means	Fuzzy Tsukamoto
Dasar Teori	Clustering unsupervised berbasis jarak Euclidean.	Fuzzy inference sederhana berbasis aturan linguistik.
Parameter awal	Centroid awal (75, 85, 95).	Aturan batas nilai (60, 79, 80).
Hasil cluster	Bisa "menyesuaikan" centroid dengan data sebenarnya, misal cluster Sedang rata-rata di sekitar 84–85.	Hasil kategorinya fix sesuai aturan, semua nilai ≥80 otomatis Tinggi.
Distribusi siswa	Lebih seimbang (ada yang masuk Sedang meski nilainya 84–85, karena lebih dekat ke centroid 85).	Hampir semua siswa masuk "Tinggi" karena rata-rata nilai di dataset di atas 80.
Kepekaan data	Fleksibel: jika ada banyak siswa nilainya 70–75, centroid rendah akan bergeser.	Kaku: tetap cut-off 60–79 & ≥80 meski distribusi data tidak seimbang.
Output	Cluster ID + kategori + centroid final.	Kategori linguistik + bobot diskrit (0, 5, 10).

Perbandingan Hasil (3 siswa)

Nama Siswa	Rata-rata	K-Means (cluster)	Fuzzy Tsukamoto
Abdul Rayhan	84.1	Sedang (C2 dekat 85)	Tinggi (≥80 → 10)
Aira Amelia	85.9	Sedang (C2 ~85)	Tinggi (≥80 → 10)

Chirul Rifai	84.8	Sedang (C2 ~85)	Tinggi (≥80 → 10)

K-Means bisa menempatkan siswa dengan nilai sekitar 84–86 sebagai Sedang, karena centroid Tinggi (95) terlalu jauh. Fuzzy Tsukamoto langsung mengkategorikan semua nilai ≥80 sebagai Tinggi, tanpa mempertimbangkan distribusi relatif antar siswa. Jadi kesimpulannya Jika tujuan mengelompokkan relatif terhadap teman sekelas K-Means lebih cocok, karena bisa membedakan "Sedang vs Tinggi" meski semua nilainya tinggi. Jika tujuanmu klasifikasi mutlak berdasar standar nilai Fuzzy Tsukamoto lebih cocok, karena aturan sudah jelas (cut-off ≥80 = Tinggi).

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis, perhitungan manual, dan implementasi program menggunakan algoritma K-Means dan Fuzzy Tsukamoto, dapat disimpulkan beberapa hal berikut:

1. Algoritma K-Means berhasil mengelompokkan siswa ke dalam tiga kategori prestasi (Rendah, Sedang, Tinggi) dengan cara membandingkan kedekatan rata-rata nilai siswa terhadap centroid awal yang ditentukan (75, 85, 95). Hasil pengelompokan menunjukkan bahwa K-Means mampu memberikan pembagian yang lebih proporsional dan relatif terhadap distribusi data siswa, sehingga siswa dengan nilai rata-rata di sekitar 84–86 dapat masuk ke cluster "Sedang" karena lebih dekat ke centroid cluster tersebut.
2. Algoritma Fuzzy Tsukamoto dengan aturan sederhana (0 untuk Rendah, 5 untuk Sedang, dan 10 untuk Tinggi) menghasilkan klasifikasi berdasarkan batas nilai mutlak. Dari hasil pengujian, sebagian besar siswa dengan rata-rata nilai di atas 80 secara otomatis dikategorikan sebagai "Tinggi" dengan bobot fuzzy 10. Hal ini membuat distribusi kategori menjadi lebih berat ke kelas "Tinggi" karena mayoritas siswa memiliki rata-rata nilai ≥80.
3. Perbandingan hasil menunjukkan bahwa K-Means lebih adaptif terhadap distribusi data karena kategori ditentukan relatif terhadap seluruh populasi, sedangkan Fuzzy Tsukamoto lebih kaku karena hanya menggunakan aturan batas nilai tetap. Dengan demikian, K-Means lebih sesuai untuk pemetaan tingkat prestasi siswa secara relatif antarindividu, sementara Fuzzy Tsukamoto

lebih sesuai untuk penilaian klasifikasi mutlak berdasarkan standar akademik tertentu.

4. Dari sisi implementasi, kedua metode dapat diterapkan dengan mudah pada dataset nilai siswa yang tersedia dalam format Excel. Program Python yang dibuat mampu menghitung rata-rata nilai per mata pelajaran, rata-rata keseluruhan, serta memberikan kategori prestasi secara otomatis sesuai dengan algoritma yang digunakan.

5.2 Saran

Berdasarkan hasil penelitian dan implementasi sistem pengelompokan prestasi siswa menggunakan algoritma K-Means dan Fuzzy Tsukamoto, maka saran yang dapat diberikan adalah sebagai berikut:

1. Penggunaan K-Means sebaiknya dilakukan apabila tujuan utama adalah melihat pemetaan prestasi siswa secara relatif di dalam satu kelas atau kelompok, sehingga guru atau pihak sekolah dapat mengetahui perbedaan tingkat capaian siswa secara lebih proporsional.
2. Penggunaan Fuzzy Tsukamoto lebih tepat apabila sekolah ingin menerapkan standar baku penilaian berdasarkan rentang nilai tertentu (misalnya nilai rata-rata ≥80 langsung dikategorikan sebagai Tinggi). Dengan demikian, metode ini cocok untuk keperluan evaluasi berdasarkan kriteria nilai absolut.
3. Disarankan agar penelitian selanjutnya juga menguji algoritma lain (misalnya Fuzzy C-Means, Naïve Bayes, atau Decision Tree) untuk dibandingkan dengan K-Means dan Fuzzy Tsukamoto, sehingga dapat diketahui metode mana yang paling sesuai dengan karakteristik data nilai siswa.

DAFTAR PUSTAKA

Adawiyah, Q., Defit, S., & Sumijan. (2024). Penerapan Algoritma K-Means Clustering untuk Mengelompokkan Rekomendasi Metode Kontrasepsi Berbasis Machine Learning di Puskesmas. *Jurnal KomtekInfo*, 300–305. https://doi.org/10.35134/komtekinfo.v11i4.563

Adzika, A. K., & Djagba, P. (2024). *Inference with K-means*. http://arxiv.org/abs/2410.17256

Anastassia, S., Kharis, A., Haqqi, A., & Zili, A. (2024). Prosiding Seminar Nasional Sains dan Teknologi Seri 02 Fakultas Sains dan Teknologi. In *Universitas Terbuka* (Vol. 1, Issue 2).

Ardianti, M., Nurhayati, O. D., & Warsito, B. (2024). Model Prediksi Kinerja Siswa Berdasarkan Data Log LMS Menggunakan Ensemble Machine Learning. *JST (Jurnal Sains Dan Teknologi)*, *12*(3). https://doi.org/10.23887/jstundiksha.v12i3.59816

Ariesanto Akhmad, E. P. (2020). Data Mining Menggunakan Regresi Linear untuk Prediksi Harga Saham Perusahaan Pelayaran. *Jurnal Aplikasi Pelayaran Dan Kepelabuhanan*, *10*(2), 120. https://doi.org/10.30649/japk.v10i2.83

Ayu, D., Wulandari, N., & Prasetyo, A. (2018). Sistem Penunjang Keputusan Untuk Menentukan Status Gizi Balita Menggunakan Metode Fuzzy Tsukamoto. *JURNAL INFORMATIKA*, *5*(1), 22–33. www.depkes.co.id

Eka Agustin, F. M., Fitria, A., & Hanifah, A. S. (2015). *IMPLEMENTASI ALGORITMA K-MEANS UNTUK MENENTUKAN KELOMPOK PENGAYAAN MATERI MATA PELAJARAN UJIAN NASIONAL (STUDI KASUS: SMP NEGERI 101 JAKARTA)* (Vol. 8, Issue 1).

Elisna, F., Zunaidi, M., & Erwansyah, K. (n.d.). Penerapan Data Mining Untuk Menganalisa Tingkat Kepuasan Pelanggan Telkomsel Terhadap Sikap Pelayanan Caroline Officer Dengan Metode Menggunakan Algoritma K-Means Clustering. *Jurnal CyberTech*, *x. No.x*.

Fernando Ade Pratama, E., & Jumadi, J. (n.d.). Kampus I: Jl Meranti Raya No.32 Sawah Lebar Kota Bengkulu 38228 Telp. (0736) 22027, Fax. *Jurnal Media Infotama*, *18*(2), 341139.

Gupta, N. S., SAgrawal, B., Chauhan, R. M., & Smt Mehta, A. B. (2015). Survey on Clustering Techniques of Data Mining. *American International Journal of Research in Science*, 15–188. http://www.iasir.net

Karlina, L., & Nurdiawan, O. (2023). PENERAPAN K-MEDOIDS DALAM KLASIFIKASI PERSEBARAN LAHAN KRITIS DI JAWA BARAT BERDASARKAN KABUPATEN/KOTA. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 1). https://opendata.jabarprov.go.id/id.

Liu, Y., Ma, S., & Du, X. (2024). A Novel Effective Distance Measure and a Relevant Algorithm for Optimizing the Initial Cluster Centroids of K-means. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.3044069

Marcelina, D., Kurnia, A., & Terttiaavini, T. (2023). Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *3*(2), 293–301. https://doi.org/10.57152/malcom.v3i2.952

Nasution, Y. R., & Eka, M. (2018). PENERAPAN ALGORITMA K-MEANS CLUSTERING PADA APLIKASI MENENTUKAN BERAT BADAN IDEAL. In *ALGORITMA: Jurnal Ilmu Komputer dan Informatika*.

Nur, F., Fauzan, R., Aziz, J., Darma Setiawan, B., & Arwani, I. (2018). *Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa* (Vol. 2, Issue 6). http://j-ptiik.ub.ac.id

Nurul Fatma Dewi Mardianto, & Yahfizham Yahfizham. (2024). Systematic Literature Review: Penerapan Berpikir Komputasi Dalam Pembelajaran Matematika. *Journal of Student Research*, *2*(4), 41–55. https://doi.org/10.55606/jsr.v2i4.3082

Oktario Dacwanda, D., & Nataliani, Y. (2021). Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan. *AITI: Jurnal Teknologi Informasi*, *18*(Agustus), 125–138.

Papakyriakou, D., & Barbounakis, I. S. (2022). Data Mining Methods: A Review. In *International Journal of Computer Applications* (Vol. 183, Issue 48).

Priya P, & Souza, D. A. D. (n.d.). *Analysis of K-Means Clustering Based Image Segmentation*. *10*(2), PP. https://doi.org/10.9790/2834-10210106

Priyatman, H., Sajid, F., & Haldivany, D. (2019). *JEPIN (Jurnal Edukasi dan Penelitian Informatika) Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa*.

Pujiarso Nugroho, R., Darma Setiawan, B., & Tanzil Furqon, M. (2019). *Penerapan Metode Fuzzy Tsukamoto untuk Menentukan Harga Sewa Hotel (Studi Kasus: Gili Amor Boutique Resort, Dusun Gili Trawangan, Nusa Tenggara Barat)* (Vol. 3, Issue 3). http://j-ptiik.ub.ac.id

Putra Eriansya, M. I., & Syafrullah, M. (2018). *Implementasi Algoritma ST-DBSCAN dan K-MEANS Untuk Pengelompokan Indeks Pembangunan Manusia Kabupaten/Kota Pulau Jawa Tahun 2014-2016 Berbasis Web Di Badan Pusat Statistik* (Vol. 1, Issue 3).

Putra, R. R., & Wadisman, C. (2018). IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL MENGGUNAKAN ALGORITMA K-MEANS IMPLEMENTATION OF DATA MINING FOR POTENTIAL CUSTOMER SELECTION USING K-MEANS ALGORITHM. *Journal of Information Technology and Computer Science (INTECOMS)*, *1*(1).

Putri, D., Asih, A., Irawan, B., & Bahtiar, A. (2024). PENGELOMPOKAN DATA TRANSAKSI DALAM MENENTUKAN STRATEGI PENJUALAN MENGGUNAKAN ALGORITMA K-MEANS. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 1).

Rezky, M. P., Sutarto, J., Prihatin, T., Yulianto, A., Haidar, I., & Surel, A. (2019). *To cite this article: Prosiding Seminar Nasional Pascasarjana UNNES*.

Rizal, S. (n.d.). Development of Big Data Analytics Model. *ITEJ Juli-2019*, *4*(1).

Sabrina Aulia Rahmah. (n.d.). *23. Sementara itu, Xu dan Wunsch*.

Saligkaras, D., & Papageorgiou, V. E. (n.d.). *Seeking the Truth Beyond the Data. An Unsupervised Machine Learning Approach*.

Sapriadi, S., Hayati, N., Eko Syaputra, A., Septi Eirlangga, Y., Manurung, K. H., & Hayati, N. (2023). Sistem Pakar Diagnosa Gaya Belajar Mahasiswa Menggunakan Metode Forward Chaining. *Jurnal Informasi Dan Teknologi*, *5*(3), 71–78. https://doi.org/10.60083/jidt.v5i3.381

Yuli Mardi. (n.d.). *Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5 Yuli Mardi*.

Yunita, F. (2018). PENERAPAN DATA MINING MENGGUNKAN ALGORITMA K-MEANS CLUSTRING PADA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS ISLAM INDRAGIRI). In *Jurnal SISTEMASI* (Vol. 7).

Zuhal, N. K. (2022). Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering. In *Universitas Nusantara PGRI Kediri. Kediri* (Vol. 1).

Lampiran

Source code K-means

```
import re
import numpy as np
import pandas as pd
from pathlib import Path

EXCEL_FILE  = "datasekolah.xlsx"
EXCEL_SHEET = 0
OUTPUT_FILE = "hasil_cluster.xlsx"

# ---------- util ----------
def _norm(s: str) -> str:
    s = (str(s) if s is not None else "").strip()
    s = s.replace("\n", " ").replace("\r", " ")
    return " ".join(s.split())

def _lower(s: str) -> str:
    return _norm(s).lower()

def make_unique(names):
    seen = {}
    out = []
    for n in names:
        if n not in seen:
            seen[n] = 0
            out.append(n)
        else:
            seen[n] += 1
            out.append(f"{n}.{seen[n]}")
    return out

def clean_columns_simple(cols):
    newcols = []
    for c in cols:
        s = str(c)
        s = re.sub(r"\s*nan\s*$", "", s, flags=re.IGNORECASE)
        s = re.sub(r"^nan\s*", "", s, flags=re.IGNORECASE)
        newcols.append(_norm(s))
    return make_unique(newcols)

# ---------- rekonstruksi kolom dari 3 baris header ----------
def reconstruct_columns_three_rows(path, sheet):
    raw = pd.read_excel(path, sheet_name=sheet, header=None, dtype=str)
    nrow, ncol = raw.shape
```

```
# cari baris identitas (mengandung 'NAMA SISWA')
id_row = None
for i in range(min(nrow, 200)):
    joined = " ".join([str(x) if pd.notna(x) else "" for x in raw.iloc[i].tolist()])
    if re.search(r"nama\s*siswa", joined, re.IGNORECASE):
        id_row = i
        break
if id_row is None:
    raise RuntimeError("Tidak menemukan baris 'NAMA SISWA'.")

mapel_row = id_row + 1
sem_row   = id_row + 2
if sem_row >= nrow:
    raise RuntimeError("Tidak cukup baris untuk MAPEL dan SEMESTER.")

id_vals   = raw.iloc[id_row].astype(str).fillna("").tolist()
mapel_val = raw.iloc[mapel_row].astype(str).fillna("").tolist()
sem_val   = raw.iloc[sem_row].astype(str).fillna("").tolist()

# forward-fill mapel
ff_mapel, last = [], ""
for j in range(ncol):
    cell = _norm(mapel_val[j])
    if cell == "" or cell.lower().startswith("unnamed") or cell.lower() == "nan":
        ff_mapel.append(last)
    else:
        ff_mapel.append(cell)
        last = cell

SEM_SET = {"I","II","III","IV"}

# bangun kolom awal
cols = []
for j in range(ncol):
    subj, sem, idn = _norm(ff_mapel[j]), _norm(sem_val[j]), _norm(id_vals[j])
    if sem in SEM_SET:
        cols.append(f"{subj} {sem}" if subj else sem)
    else:
        cols.append(idn if idn else subj)

# tempelkan mapel ke kolom "II/III/IV"
fixed, current_subject = [], ""
sem_regex = re.compile(r"^(I{1,3}|IV)(?:\.\d+)?$", re.IGNORECASE)
for name in cols:
    n = _norm(name)
    m_pair = re.match(r"^(.*)\s+(I{1,3}|IV)(?:\.\d+)?$", n, flags=re.IGNORECASE)
    if m_pair:
        current_subject = _norm(m_pair.group(1))
        fixed.append(f"{current_subject} {m_pair.group(2).upper()}")
```

```
            continue
        if sem_regex.match(n):
            fixed.append(f"{current_subject} {n.split('.')[0].upper()}" if current_subject else n.upper())
            continue
        fixed.append(n)
    fixed = clean_columns_simple(fixed)

    df = raw.iloc[sem_row+1:].copy()
    df.columns = fixed
    df = df.dropna(how="all").reset_index(drop=True)
    return df

# ---------- deteksi kolom nama ----------
def detect_name_column(df):
    for c in df.columns:
        if re.fullmatch(r"(?i)nama\s*siswa", c):
            return c
    for c in df.columns:
        if "nama" in c.lower() and "nis" not in c.lower():
            return c
    return df.columns[0]

# ---------- mapping mapel ----------
MAPEL_NAMES = ["PAI","PPKN","B.INDONESIA","MATEMATIKA","IPA","IPS","B.
INGGRIS","SENI BUDAYA","PJOK","PRAKARYA","IQRA","TIK"]
SEM_LABELS = ["I","II","III","IV"]

def build_expected():
    exp = {}
    for m in MAPEL_NAMES:
        vars = [m, m.replace(". ","."), m.replace(".",""), m.replace("  "," "), m.replace("."," .")]
        vars = list(dict.fromkeys(vars))
        names=[]
        for sem in SEM_LABELS:
            names.append((sem,[f"{v} {sem}" for v in vars]))
        exp[m]=names
    return exp

EXPECTED = build_expected()

def find_semester_cols(df, mapel):
    cols=[]
    low={_lower(c):c for c in df.columns}
    for sem,cands in EXPECTED[mapel]:
        found=None
        for cand in cands:
            key=_lower(cand)
            if key in low:
                found=low[key]; break
```

```
        if found: cols.append(found)
    return cols

# ---------- kmeans ----------
def kmeans_1d(values, centroids=(75,85,95), max_iter=100, tol=1e-6):
    centroids=list(map(float,centroids))
    x=np.asarray(values,float)
    for _ in range(max_iter):
        d=np.abs(x.reshape(-1,1)-np.asarray(centroids).reshape(1,-1))
        labels=d.argmin(1)
        new=[]
        for k in range(len(centroids)):
            mem=x[labels==k]
            new.append(float(mem.mean()) if len(mem) else centroids[k])
        if max(abs(a-b) for a,b in zip(new,centroids))<=tol:
            centroids=new; break
        centroids=new
    return labels,centroids

# ---------- main ----------
def main():
    if not Path(EXCEL_FILE).exists():
        raise FileNotFoundError(EXCEL_FILE)

    df = reconstruct_columns_three_rows(EXCEL_FILE, EXCEL_SHEET)

    print("\n[DEBUG] Kolom (awal):", df.columns[:60].tolist())
    col_nama = detect_name_column(df)
    print("[INFO] Kolom nama:", col_nama)

    # numerikkan semua kecuali nama
    for c in df.columns:
        if c != col_nama:
            df[c] = pd.to_numeric(df[c], errors="coerce")

    # rata-rata per mapel
    avg_cols = []
    for m in MAPEL_NAMES:
        cols = find_semester_cols(df, m)
        if len(cols) == 4:
            df[f"{m}_AVG"] = df[cols].mean(axis=1, skipna=True)
            avg_cols.append(f"{m}_AVG")
        else:
            print(f"[PERINGATAN] {m}: hanya {len(cols)} dari 4 kolom ditemukan -> dilewati")

    if not avg_cols:
        raise RuntimeError("Tidak ada mapel terpetakan (cek header).")

    # rata-rata total
```

```
    df["RATA_RATA"] = df[avg_cols].mean(axis=1, skipna=True)

    # --- valid vs tidak lengkap ---
    valid   = df[avg_cols].notna().any(axis=1)   # <- FIX axis=1
    invalid = ~valid
    df.loc[invalid, "KATEGORI"] = "Data tidak lengkap"

    # k-means hanya pada baris valid
    if valid.any():
        vals = df.loc[valid, "RATA_RATA"].values
        labels, cent = kmeans_1d(vals, (75,85,95))
        order = np.argsort(cent)
        label_map = {order[0]:"Rendah", order[1]:"Sedang", order[2]:"Tinggi"}

        df.loc[valid, "CLUSTER_ID"] = labels
        df.loc[valid, "KATEGORI"]   = df.loc[valid, "CLUSTER_ID"].map(label_map)

        print("\nCentroid akhir:", [round(c,2) for c in cent])
        counts = df.loc[valid, "KATEGORI"].value_counts().to_dict()
        print("Jumlah anggota per cluster (valid):", counts)
    else:
        print("\n[TIDAK ADA DATA VALID] Semua baris 'Data tidak lengkap'.")

    # output ringkas
    ringkas = df[[col_nama, "RATA_RATA", "KATEGORI"]].copy().sort_values("RATA_RATA",
ascending=False)
    print("\n== Hasil Ringkas ==")
    with pd.option_context("display.max_rows", None, "display.width", 180):
        print(ringkas.to_string(index=False))

    # simpan excel
    with pd.ExcelWriter(OUTPUT_FILE, engine="xlsxwriter") as w:
        ringkas.to_excel(w, index=False, sheet_name="Cluster")
        rincian = df[[col_nama] + avg_cols + ["RATA_RATA", "KATEGORI"]]
        rincian.to_excel(w, index=False, sheet_name="Rincian")

    print("\n[OK] Disimpan ke:", OUTPUT_FILE)

if __name__=="__main__":
    main()
```

Sourcode Fuzzy Tsukamoto

```
# fuzzy_tsukamoto.py
# ------------------------------------------------------
# Fuzzy Tsukamoto  untuk bobot prestasi siswa.
# - Baca Excel "datasekolah.xlsx" (header merge 3 baris: identitas / MAPEL / I..IV)
```

```
# - Rata-rata per mapel (dari 4 semester), lalu rata-rata 12 mapel per siswa
# - Aturan fuzzy (diskrit): x<60->0 (Rendah), 60<=x<79->5 (Sedang), x>=80->10 (Tinggi)
# - Output: hasil_fuzzy_tsukamoto.xlsx
# ------------------------------------------------------------

import re
from pathlib import Path
import numpy as np
import pandas as pd

EXCEL_FILE  = "datasekolah.xlsx"
EXCEL_SHEET = 0
OUTPUT_FILE = "hasil_fuzzy_tsukamoto.xlsx"

MAPEL_NAMES = [
    "PAI","PPKN","B.INDONESIA","MATEMATIKA","IPA","IPS",
    "B. INGGRIS","SENI BUDAYA","PJOK","PRAKARYA","IQRA","TIK"
]
SEM_LABELS = ["I","II","III","IV"]

# ========== util ==========
def _norm(s: str) -> str:
    s = (str(s) if s is not None else "").strip()
    s = s.replace("\n"," ").replace("\r"," ")
    return " ".join(s.split())

def _lower(s: str) -> str:
    return _norm(s).lower()

def make_unique(names):
    seen, out = {}, []
    for n in names:
        if n not in seen:
            seen[n] = 0
            out.append(n)
        else:
            seen[n] += 1
            out.append(f"{n}.{seen[n]}")
    return out

def clean_columns_simple(cols):
    newcols = []
    for c in cols:
        s = str(c)
        s = re.sub(r"\s*nan\s*$","", s, flags=re.IGNORECASE)
        s = re.sub(r"^nan\s*","", s, flags=re.IGNORECASE)
        newcols.append(_norm(s))
    return make_unique(newcols)
```

```
# ========== rekonstruksi header 3 baris ==========
def reconstruct_columns_three_rows(path, sheet):
    raw = pd.read_excel(path, sheet_name=sheet, header=None, dtype=str)
    nrow, ncol = raw.shape

    # Cari baris yang memuat "NAMA SISWA"
    id_row = None
    for i in range(min(200, nrow)):
        joined = " ".join([str(x) if pd.notna(x) else "" for x in raw.iloc[i].tolist()])
        if re.search(r"nama\s*siswa", joined, re.IGNORECASE):
            id_row = i
            break
    if id_row is None:
        raise RuntimeError("Tidak menemukan baris 'NAMA SISWA' di Excel.")

    mapel_row = id_row + 1
    sem_row   = id_row + 2
    if sem_row >= nrow:
        raise RuntimeError("Format header tidak lengkap (butuh 3 baris).")

    id_vals   = raw.iloc[id_row].astype(str).fillna("").tolist()
    mapel_val = raw.iloc[mapel_row].astype(str).fillna("").tolist()
    sem_val   = raw.iloc[sem_row].astype(str).fillna("").tolist()

    # Forward-fill nama mapel
    ff_mapel, last = [], ""
    for j in range(ncol):
        cell = _norm(mapel_val[j])
        if cell == "" or cell.lower().startswith("unnamed") or cell.lower() == "nan":
            ff_mapel.append(last)
        else:
            ff_mapel.append(cell); last = cell

    SEM_SET = {"I","II","III","IV"}
    cols = []
    for j in range(ncol):
        subj, sem, idn = _norm(ff_mapel[j]), _norm(sem_val[j]), _norm(id_vals[j])
        if sem in SEM_SET:
            cols.append(f"{subj} {sem}" if subj else sem)
        else:
            cols.append(idn if idn else subj)

    # “Tempelkan” mapel ke kolom yang hanya II/III/IV
    fixed, current_subject = [], ""
    sem_regex = re.compile(r"^(I{1,3}|IV)(?:\.\d+)?$", re.IGNORECASE)
    for name in cols:
        n = _norm(name)
        m_pair = re.match(r"^(.*)\s+(I{1,3}|IV)(?:\.\d+)?$", n, flags=re.IGNORECASE)
        if m_pair:
```

```
            current_subject = _norm(m_pair.group(1))
            fixed.append(f"{current_subject} {m_pair.group(2).upper()}")
            continue
        if sem_regex.match(n):
            fixed.append(f"{current_subject} {n.split('.')[0].upper()}" if current_subject else n.upper())
            continue
        fixed.append(n)
    fixed = clean_columns_simple(fixed)

    df = raw.iloc[sem_row+1:].copy()
    df.columns = fixed
    df = df.dropna(how="all").reset_index(drop=True)
    return df

# ========== deteksi kolom nama siswa ==========
def detect_name_column(df: pd.DataFrame) -> str:
    for c in df.columns:
        if re.fullmatch(r"(?i)nama\s*siswa", c):
            return c
    for c in df.columns:
        if "nama" in c.lower() and "nis" not in c.lower():
            return c
    # fallback
    return df.columns[0]

# ========== mapping nama kolom mapel I..IV yang fleksibel ==========
def build_expected():
    exp = {}
    for m in MAPEL_NAMES:
        variants = [
            m, m.replace(". ", "."), m.replace(".", ""), m.replace("  ", " "),
            m.replace("."," .")  # toleransi kecil
        ]
        variants = list(dict.fromkeys(variants))
        pairs = []
        for sem in SEM_LABELS:
            pairs.append((sem, [f"{v} {sem}" for v in variants]))
        exp[m] = pairs
    return exp

EXPECTED = build_expected()

def find_semester_cols(df: pd.DataFrame, mapel: str):
    """Cari 4 kolom untuk mapel tertentu (I..IV), toleransi variasi spasi/titik."""
    cols = []
    low = {_lower(c): c for c in df.columns}
    for sem, candidates in EXPECTED[mapel]:
        found = None
        for cand in candidates:
```

```
            key = _lower(cand)
            if key in low:
                found = low[key]; break
        if found: cols.append(found)
    return cols  # urut sesuai SEM_LABELS

# ========== fuzzy Tsukamoto (diskrit) ==========
def fuzzy_bobot(x: float):
    """Aturan diskrit sesuai permintaan:
       x < 60 -> (0, 'Rendah')
       60 <= x < 79 -> (5, 'Sedang')
       x >= 80 -> (10, 'Tinggi')
    """
    if pd.isna(x):
        return np.nan, "Data tidak lengkap"
    if x < 60:
        return 0, "Rendah"
    elif x < 79:
        return 5, "Sedang"
    else:
        return 10, "Tinggi"

# ========== main ==========
def main():
    if not Path(EXCEL_FILE).exists():
        raise FileNotFoundError(f"Tidak menemukan file {EXCEL_FILE}")

    df = reconstruct_columns_three_rows(EXCEL_FILE, EXCEL_SHEET)

    # Info awal
    print("\n[DEBUG] 50 kolom pertama:", df.columns[:50].tolist())

    # Deteksi kolom nama
    col_nama = detect_name_column(df)
    print("[INFO] Kolom nama terdeteksi:", col_nama)

    # Konversi semua (kecuali nama) ke numerik
    for c in df.columns:
        if c != col_nama:
            df[c] = pd.to_numeric(df[c], errors="coerce")

    # Hitung rata-rata per mapel
    avg_cols = []
    for m in MAPEL_NAMES:
        cols = find_semester_cols(df, m)
        if len(cols) == 4:
            df[f"{m}_AVG"] = df[cols].mean(axis=1, skipna=True)
            avg_cols.append(f"{m}_AVG")
        else:
```

```
            print(f"[PERINGATAN] Kolom {m} tidak lengkap (ketemu {len(cols)}/4) — dilewati.")

    if not avg_cols:
        raise RuntimeError("Tidak ada mapel yang berhasil dipetakan.")

    # Rata-rata keseluruhan 12 mapel (yang tersedia)
    df["RATA_RATA"] = df[avg_cols].mean(axis=1, skipna=True)

    # Fuzzy bobot + kategori
    bobot_list, kategori_list = [], []
    for v in df["RATA_RATA"].tolist():
        b, k = fuzzy_bobot(v)
        bobot_list.append(b); kategori_list.append(k)
    df["BOBOT"] = bobot_list
    df["KATEGORI"] = kategori_list

    # Ringkas
    ringkas = df[[col_nama, "RATA_RATA", "KATEGORI", "BOBOT"]].copy() \
                .sort_values("RATA_RATA", ascending=False)

    # Cetak khusus: Chirul Rifai (sesuai contoh)
    mask_chirul = df[col_nama].astype(str).str.contains(r"\bChirul\s+Rifai\b", case=False,
regex=True)
    if mask_chirul.any():
        row = df.loc[mask_chirul].iloc[0]
        print("\n[INFO] Contoh (Chirul Rifai)")
        print(f"Rata-rata per mapel (12): {', '.join([f'{row[c]:.2f}' for c in avg_cols if not
pd.isna(row[c])])}")
        print(f"Rata-rata keseluruhan: {row['RATA_RATA']:.2f}")
        print(f"Kategori: {row['KATEGORI']} | Bobot: {int(row['BOBOT']) if not pd.isna(row['BOBOT'])
else 'NaN'}")

    # Tampilkan ringkasan ke layar
    print("\n== Hasil Ringkas (top 20) ==")
    with pd.option_context("display.max_rows", 20, "display.width", 180):
        print(ringkas.to_string(index=False))

    # Simpan Excel
    with pd.ExcelWriter(OUTPUT_FILE, engine="xlsxwriter") as w:
        ringkas.to_excel(w, index=False, sheet_name="Ringkas")
        rincian = df[[col_nama] + avg_cols + ["RATA_RATA", "KATEGORI", "BOBOT"]]
        rincian.to_excel(w, index=False, sheet_name="Rincian")

    print(f"\n[OK] Hasil disimpan ke: {OUTPUT_FILE}")

if __name__ == "__main__":
    main()
```