

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

2.1.1. Kinerja Siswa

Salah satu elemen penting yang sangat diperlukan dan dilakukan di bidang bisnis dan pendidikan adalah pengukuran kinerja. Dengan tujuan untuk terus meningkatkan kualitas pendidikan, Direktorat Jenderal Pendidikan Tinggi telah memasukkan pengukuran kinerja ke dalam dunia pendidikan atau pengelolaan perguruan tinggi (Nuryadin, 2013). Penilaian, akreditasi, dan evaluasi diri digunakan untuk meningkatkan kualitas pendidikan melalui peningkatan kompetensi siswa.

Kemampuan untuk melakukan tugas tertentu yang didasarkan pada pengetahuan, keterampilan serta didukung dengan sikap kerja yang mumpuni oleh posisi tersebut disebut kompetensi (Posuma, 2013). Penentuan kompetensi siswa sangat penting karena akan membantu mereka memasuki dunia kerja yang sesuai dengan kemampuan mereka. Dengan kompetensi yang dimilikinya, siswa akan memiliki landasan untuk bekerja sebagai karyawan atau pengusaha (Hadi et al., 2016).

Sekolah harus terus meningkatkan status kualitas internalnya untuk meningkatkan kualitas siswanya. Untuk mencapai tujuan ini, sekolah harus secara rutin melakukan pengawasan dan evaluasi kinerja siswa. Namun, ada kendala yang muncul seiring berjalannya waktu. Selama ini, pengolahan data hanya dilakukan menggunakan Microsoft Excel berdasarkan penilaian guru. Tidak banyak perangkat lunak yang dimaksudkan untuk mengolah data penilaian untuk melacak dan menilai kinerja siswa. Oleh karena itu, untuk melihat sejauh mana kinerja dari siswa yang ada di sekolah diperlukan sebuah penilaian terkomputerisasi melalui sebuah sistem untuk membantu pengambil keputusan dengan baik dan tepat dalam memustuskan kinerja siswa (Warasto, 2016).

Permasalahan di dalam evaluasi kinerja siswa salah satunya adalah pengelompokan, di mana kelompok dibentuk harusnya berdasarkan tingkat kecerdasan siswa dan ukuran kelas tidak boleh melebihi kapasitas yang telah ditentukan. Pengelompokan siswa berdasarkan tingkat kecerdasan sangat penting untuk menjaga homogenitas kelompok. Kelompok siswa yang tidak homogen akan sulit untuk memberikan layanan pendidikan yang baik kepada populasi yang sangat beragam. Selain itu, pengelompokan yang homogen dengan kemampuan yang sama akan lebih jauh membuat hasil kinerja siswa lebih adil, realistis, dan sebanding (Yadav & Singh, 2012). Salah satu teknik di dalam pengelompokan tersebut dapat dilakukan dengan Data Mining yang telah banyak digunakan di dalam dunia bisnis, kesehatan, dan pendidikan.

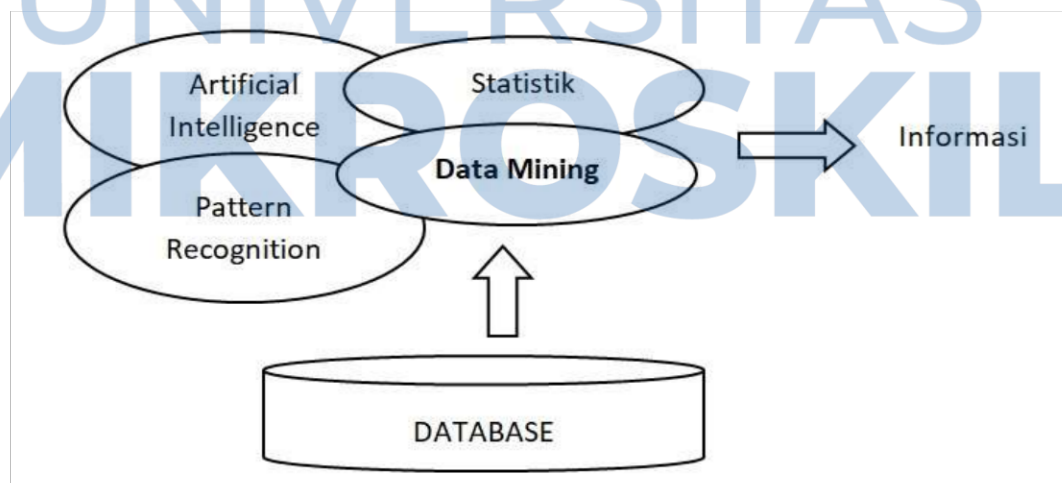
Dalam pendidikan, data mining difungsikan sebagai penganalisis data yang merupakan hasil dari proses pembelajaran untuk mendapatkan informasi yang baru dari data yang sudah pernah ada, sehingga menghasilkan peningkatan kualitas pembelajaran. Pengelompokan siswa berdasarkan potensi dapat membantu meningkatkan prestasi belajar, tingkat kelulusan, dan pengelolaan data siswa yang lebih baik (Wulandari et al., 2020). Kontribusi data mining dalam pendidikan telah digunakan untuk meningkatkan pemahaman tentang proses pendidikan dengan tujuan utama memberikan rekomendasi kepada pengajar dan peneliti untuk perbaikan proses belajar-mengajar. Penerapan aplikasi data mining dalam pendidikan membuat penyelenggara pembelajaran dapat mengatur sumber daya pendidikan dengan cara yang lebih efisien (Santoso & Yulia, 2019).

2.1.2. Data Mining

Istilah "data mining" mengacu pada teknik yang digunakan untuk menggali atau menambang pengetahuan yang masih belum terlihat atau tersembunyi dalam sekumpulan data yang banyak atau database. Teknik ini melibatkan analisis peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dari sebelumnya yang lebih mudah dipahami dan bermanfaat bagi pemilik data (Larose, 2005).

Seperti diungkapkan oleh Turban et al. (2010), istilah "data mining" difungsikan untuk menggambarkan penemuan pengetahuan yang ditemukan di dalam database. Dalam prosesnya, data mining dilakukan dengan menggunakan teknik pembelajaran mesin, statistik, kecerdasan buatan, matematika, untuk mengidentifikasi informasi dan mengetahui lebih terinci informasi penting dan relevan dari database. Kumpulan data pada sistem informasi disebut database. Teknik data mining digunakan untuk mengolah dan mempelajari data ini, menghasilkan informasi baru kemudian dianalisis dengan mudah dipahami oleh pemilik data atau orang lain.

Sesuai dengan definisi sebelumnya, bahwa data mining merupakan salah satu upaya untuk melakukan pengolahan pengetahuan lebih terinci dalam database untuk menemukan sebuah pola dengan menggunakan teknik dari berbagai keilmuan seperti kecerdasan buatan, matematika, dan statistik serta pembelajaran mesin, untuk melakukan proses identifikasi suatu informasi yang terdiri dari pengetahuan tersebut. Secara umum, data mining digunakan di berbagai bidang seperti pemasaran, keuangan, kesehatan, ilmu pengetahuan, dan lainnya untuk mengungkap wawasan yang dapat mempengaruhi strategi bisnis, meningkatkan efisiensi operasional, atau mendukung penelitian ilmiah. Adapun berbagai bidang keilmuan di dalam data mining dapat dilihat pada Gambar II-2 sebagai berikut.



Gambar II-1. Akar Ilmu Data Mining

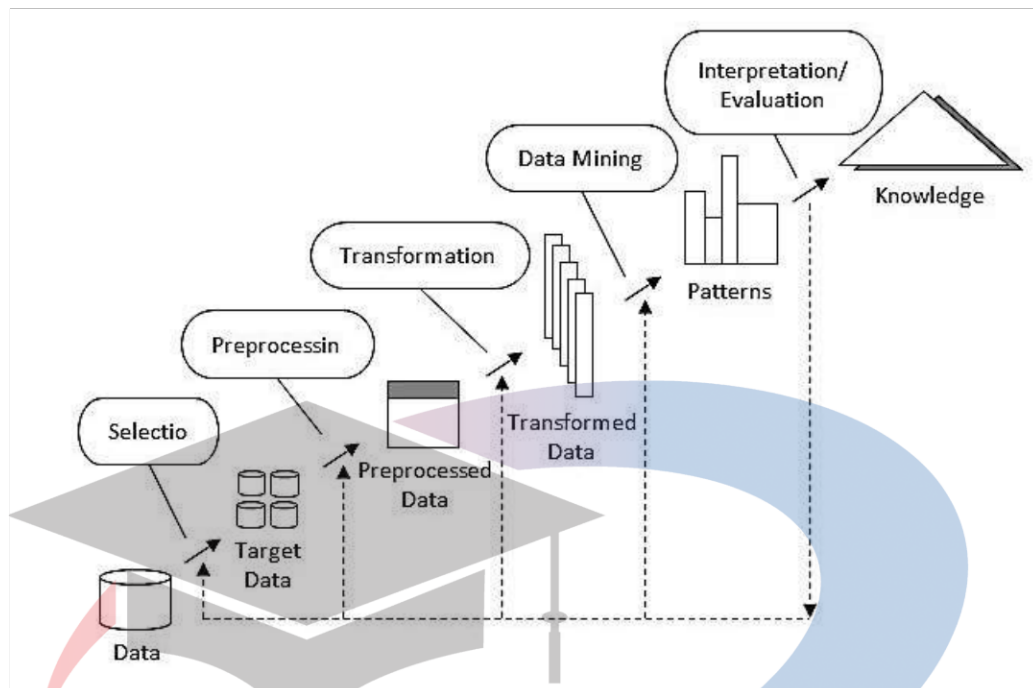
Data mining adalah proses analisis dan penggalian sejumlah besar data untuk menemukan pola tertentu dan mendapatkan kebenaran serta keuntungan.

Proses ini dilakukan berulang kali atau iterasi hingga hasilnya adalah set pola yang tepat yang memenuhi tujuan awal. Data mining terbagi menjadi dua kelompok umum. (Tan et al., 2006) sebagai berikut:

1. Proses mengumpulkan pola pada data dengan beberapa variabel untuk memprediksi variabel lain pada waktu yang akan datang dikenal sebagai *Predictive Mining*. Regresi, klasifikasi, dan deviasi adalah beberapa teknik prediktive.
2. Proses yang dilakukan untuk mendapatkan karakteristik penting pada data dalam sebuah database merupakan proses *Descriptive mining*. *Clustering, Sequential mining dan association* adalah termasuk dalam kategori teknik *Deskriptive Mining*.

Knowledge Discovery in Database (KDD) adalah proses yang dilaksanakan untuk mencari informasi yang lebih mudah dipahami, lebih baru dari penyimpanan data yang kompleks dan lebih bernilai. KDD ini merupakan istilah lain yang mempunyai makna yang serupa dengan data mining. Proses KDD menyimpulkan dalam suatu penafsiran yang merupakan hasil dari sekumpulan data dengan menggabungkan data ilmu lain. Data mining adalah salah satu bagian dari proses KDD. Dimulai dengan menetapkan tujuan, proses KDD berakhir dengan evaluasi. (Tomar & Agarwal, 2013).

Penerapan beberapa prosedur dan algoritma untuk pengambilan pola dari database yang besar merupakan Peran utama data mining. Peningkatan teknologi saat ini yang begitu pesat, begitu mudahnya data dapat diambil dari berbagai sumber dalam berbagai format seperti data ilmiah, teks, gambar, video dan format data baru. Untuk proses pengambilan keputusan yang efektif, data yang diambil dari berbagai sumber harus dianalisis dengan benar. (Sumathi et al., 2016). Tahapan dari KDD dapat dilihat pada Gambar II-2 berikut.



Gambar II-2. Tahapan Proses KDD

Fase yang terdapat pada proses *Knowledge Discovery Database* (KDD) sebagai berikut:

a. Seleksi Data (*Selection*)

Sebelum memasuki tahapan awal pada saat penggalian informasi dalam *Knowledge Discovery Database* (KDD) dimulai, terlebih dahulu data dipilih dari kumpulan data operasional harus dipilih dan dilakukan penyimpanan dalam satu dokumen yang terpisah dari database operasional.

b. Pemilihan Data (*Preprocessing/Cleaning*)

Pada saat melaksanakan preprocessing yaitu meliputi penghapusan data yang terduplikasi, tidak konsisten, dan memperbaiki kesalahan data, seperti kesalahan cetak (tipografi). Selain itu, proses pengembangan preprocessing juga melaksanakan penggabungan data yang masih berada dalam data dan data informasi lain yang relevan serta diperlukan untuk KDD, seperti informasi eksternal, dilakukan selama proses ini.

c. Transformasi (*Transformation*)

Pada fase transformasi yang dilakukan adalah melibatkan transformasi bentuk data yang tidak memiliki entitas yang jelas ke dalam bentuk data yang dapat diproses dan siap untuk diproses.

d. *Data Mining*

Menerapkan algoritma yang merupakan salah satu teknik pencarian pengetahuan adalah langkah pertama dalam proses data mining.

e. Interpretasi/Evaluasi (*Interpretation/Evaluation*)

Fase terakhir ini dilakukan melalui proses pembentukan keluaran yang mudah dipahami yang berasal dari proses data mining. Clustering, klasifikasi, asosiasi, dan outlier adalah beberapa masalah yang dihadapi oleh bisnis, dan teknik data mining dapat digunakan untuk memecahkan masalah ini. Metode dasar analisis data salah satunya adalah klasifikasi, juga disebut sebagai teknik untuk menentukan anggota kelompok dari data yang tersedia. Ide dasar dari klasifikasi adalah bahwa sejumlah data dengan struktur data yang hampir identik akan menghasilkan klasifikasi yang hampir identik atau sebanding. Metode data mining yang populer di banyak bidang adalah klasifikasi (Sung et al., 2015).

2.1.3. Pengelompokan Data

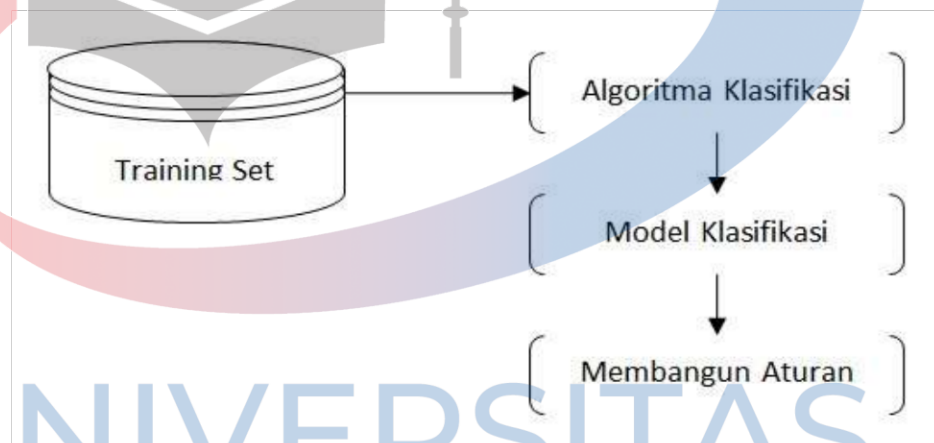
Estimation, prediction, classification, clustering, dan association adalah beberapa fungsi utama data mining. Semua fungsi data mining dibagi dalam dua metode berdasarkan metode pembelajarannya yaitu Metode pembelajaran supervised memerlukan data sampel, juga dikenal sebagai data sampel dan metode unsupervised tidak memerlukan data sampel. Klasifikasi adalah salah satu contoh peran data mining dengan metode pembelajaran supervise (Santosa, 2007).

Pada proses klasifikasi didesain untuk menemukan suatu model dan pola sehingga adanya perbedaan yang menggambarkan konsep atau class data. Tujuannya adalah agar model yang telah ditemukan dapat difungsikan untuk menghasilkan prediksi class objek yang label class-nya belum diketahui. Model didasarkan pada analisis data pelatihan, yaitu objek data dengan label kelas yang sudah diketahui. Dengan menggunakan algoritma klasifikasi untuk mempelajari

data sampel, pengelompokan data akan menemukan pola tertentu pada data sampel yang berkaitan dengan kelas target. Menurut (Han & Kamber, 2006) prediksi kelas target dapat dilakukan dengan menggunakan data di luar sampel menggunakan model klasifikasi. Klasifikasi dilakukan dalam dua tahap (Annasaheb & Verma, 2016). yaitu:

1. Tahap membangun model

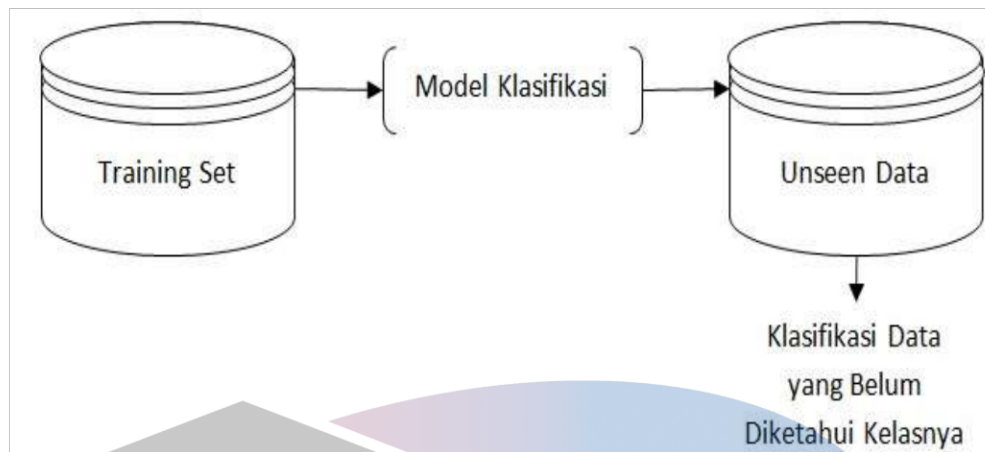
Pada tahapan mendesain pembangunan model klasifikasi didasarkan pada data kelas yang telah ditentukan. Sedangkan data sampel yang proses pada saat berlangsungnya tahapan ini disebut sebagai data pelatihan atau data pembelajaran (*training set*). Selama proses ini berlangsung disebut dengan proses induksi sesuai pada Gambar II-3 sebagai berikut.



Gambar II-3. Tahapan Membangun Model

2. Tahap menggunakan model klasifikasi

Pada tahap ini, model digunakan pada data yang kelasnya belum diketahui. Ini dilakukan dengan menggunakan data uji, atau set uji, untuk memprediksi kelas label data dalam himpunan. Proses ini disebut deduksi.. Proses ini dapat dilihat pada Gambar II-4 sebagai berikut.



Gambar II-4. Tahapan Menggunakan Model

2.1.4. Algoritma *RObust Clustering using linKs* (ROCK)

Algoritma ROCK muncul pertama kali di hadapan publik oleh Guha pada tahun 1999. Algoritma ROCK ini sebelumnya dikembangkan oleh Sugato Basu, Arindam Banerjee, and Raymond Mooney pada tahun 2003. Berdasarkan data kategori, algoritma ROCK adalah algoritma yang pengelompokan dilaksanakan hierarki agglomeratif dengan menggunakan jarak antar titik. Oleh karena itu, diciptakan ide baru, yaitu link, yang digunakan untuk mengukur kesetaraan atau jarak kedekatan dua titik data (Dewi et al., 2021).

Karakteristik utama dari algoritma ROCK adalah a. *Link Based Clustering* yaitu Algoritma ini menggunakan pendekatan berbasis link (hubungan antar titik data) daripada jarak antara titik untuk menentukan cluster. b. *Robustness* yaitu ROCK dirancang untuk lebih tahan terhadap gangguan atau noise dalam data, serta lebih stabil terhadap perubahan dalam inisialisasi cluster. c. *Parameerizable Similarity Threshold* yaitu Pengguna dapat menentukan ambang batas kesamaan (similarity threshold) untuk mengontrol bagaimana titik-titik data dihubungkan dalam proses clustering. d. *Hierarchical Clustering* yaitu Algoritma ini dapat menghasilkan struktur hierarkis dari cluster, memungkinkan analisis yang lebih mendalam terhadap hubungan antar data.

Pengelompokan data berdasarkan kategori dengan algoritma ROCK memiliki beberapa tahapan yang dilakukan sebagai berikut (Dewi et al., 2021):

1. Menghitung kemiripan (*similarity*), yaitu ukuran kemiripan antara pasangan objek ke-i dan objek ke-j dengan menggunakan Persamaan 2.1 yang didefinisikan sebagai berikut.

$$sim(x_i, x_j) = \frac{|Y_i \cap Y_j|}{|Y_i \cup Y_j|}, i \neq j \quad (2.1)$$

Keterangan:

$i = 1, 2, 3 \dots n$ dan $j = 1, 2, 3 \dots, n$

Y_i = Himpunan pengamatan ke-i, dengan

$Y_i = \{x_{1i}, x_{2i}, \dots, x_{mi}\},$

Y_j = Himpunan pengamatan ke-j, dengan

$Y_j = \{x_{1j}, x_{2j}, \dots, x_{mj}\},$

$|Y|$ = Bilangan kardinal atau jumlah anggota dari himpunan Y.

2. Menentukan tetangga dengan ketentuan, jika nilai $(x_i, x_j) \geq \theta$.
3. Perhitungan *link* antara objek yang diamati.

Menghitung link antara objek pengamatan Nilai ambang (θ), parameter yang dibuat oleh peneliti untuk mengendalikan jarak dekat link antar objek, yang dapat mempengaruhi besarnya link. Bilangan umum yang digunakan 0 dan 1.

Algoritma ROCK selesai atau berakhir pada saat jumlah kelompok yang sesuai target terpenuhi atau tidak ada lagi titik link antara kelompok satu dengan lainnya. Matriks A dapat digunakan untuk menghitung berapa jumlah link untuk keseluruhan pasangan yang mungkin dari n objek. Matriks A adalah matriks berukuran $n \times n$, dengan nilai 1 jika pasangan x_i dan x_j dinyatakan tetangga (mirip), dan nilai 0 jika pasangan x_i dan x_j dinyatakan bukan tetangga (tidak mirip). Hasil perkalian antara baris ke x_i dan kolom ke x_j di matriks A digunakan untuk menghitung jumlah link antara pasangan x_i dan x_j . Kemungkinan x_i dan x_j berada dalam kelompok yang sama meningkat, seiring dengan luasnya link antara mereka.

4. Penggabungan kelompok. Algoritma ROCK menggabungkan kelompok berdasarkan ukuran kestabilan dan konsisten (*goodness measure*) antar kelompok. Seperti rumus pada persamaan 2.2 di bawah. *Goodness*

measure adalah persamaan untuk menghitung kualitas link dan dibagi dengan kemungkinan link yang terbentuk berdasarkan ukuran kelompoknya.

$$g(C_i, C_j) = \frac{\text{link}[C_i, C_j]}{(n_i + n_j)^{1+2f(\theta)} - n_i^{1+2f(\theta)} - n_j^{1+2f(\theta)}} \quad (2.2)$$

dengan $\text{link}[C_i, C_j] = \sum_{x_i \in C_i, x_j \in C_j} \text{link}(x_i, x_j)$ yang menyatakan bahwa jumlah *link* dari semua kemungkinan pasangan objek yang ada dalam C_i dan C_j , Dimana:

n_i = jumlah anggota dalam kelompok ke- i

n_j = jumlah anggota dalam kelompok ke- j , dan

$$f(\theta) = \frac{1-\theta}{1+\theta} \quad (2.3)$$

2.1.5. Fuzzy C-Means Clustering

Clustering dalam data mining merupakan salah satu teknik komputasi yang berfungsi untuk mengidentifikasi pola yang mirip/sama pada satu atau lebih data dalam sebuah kelompok data. Data dapat dikelompokkan ke dalam kelompok berdasarkan tingkat kemiripan yang sama. Metode FCM memungkinkan tiap vektor pada seluruh atribut data dimiliki lebih dari satu cluster dengan derajat keanggotaan yang berbeda-beda. Untuk membedakan derajat keanggotaan dapat dibedakan dengan menggunakan bilangan dari 0 hingga 1.

Metode mengelompokkan data *Fuzzy C-Means Clustering* (FCM) mengukur tingkat kesamaan atau ketidaksamaan di antara pasangan kelompok data. Nilai derajat keanggotaan yang tinggi dan data yang sangat mirip satu dengan yang lainnya akan diklasifikasikan dalam kelompok yang sama. Konsep utama metode FCM adalah menentukan beberapa pusat cluster yang ditandai dengan menghitung lokasi rata-rata setiap cluster. Untuk memastikan bahwa pusat cluster akan bergerak menuju lokasi yang tepat yaitu dengan memperbaiki derajat keanggotaan dan pusat cluster setiap titik data secara berulang. Dasar dari perlakuan perulangan ini yaitu nilai minimal dari sebuah fungsi objektif, yang menunjukkan jarak dari titik data yang diberikan ke pusat cluster yang terbobot oleh derajat keanggotaan tersebut.

Metode FCM menghasilkan luaran berupa derajat keanggotaan dan deretan pusat cluster untuk tiap titik data. Beberapa langkah-langkah sebagai prosedur yang dilakukan untuk penerapan metode FCM dapat dijelaskan sebagai berikut (kusumadewi dkk, 2010) :

1. Persiapkan terlebih dahulu sejumlah data yang digunakan untuk proses *clustering*. Misalnya, matriks X merupakan data yang berukuran $i \times j$, dengan i = jumlah data yang akan di-*cluster* dan j = jumlah atribut.

$$\begin{bmatrix} X_{11} & X_{12} & \dots & X_{1j} \\ X_{21} & X_{22} & \dots & X_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ X_{i1} & X_{i2} & \dots & X_{ij} \end{bmatrix}$$

Gambar II-5. Data Matrik Berukuran $i \times j$

2. Menentukan adanya beberapa indikator atau parameter sesuai keterangan dibawah ini
 - a. Jumlah *cluster* yang akan dibentuk ($n > c \geq 2$)
 - b. Pembobot ($w > 1$)
 - c. Maksimum iterasi ($max\ n$)
 - d. Kriteria penghentian/*threshol*d (ϵ = nilai positif yang sangat kecil)
 - e. Menentukan fungsi obyektif awal (P_0)
3. Derajat keanggotaan dibentuk untuk setiap *cluster* dengan ukuran $i \times k$ terbentuk dari matriks partisi awal u , dengan i = jumlah data yang akan di-*cluster* dan k = jumlah *cluster*. Matriks partisi biasanya dibuat acak dengan rentang nilai 0 sampai dengan 1.

$$\begin{bmatrix} u_{11} & u_{12} & \dots & u_{1k} \\ u_{21} & u_{22} & \dots & u_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ u_{i1} & u_{i2} & \dots & u_{ik} \end{bmatrix}$$

Gambar II-6. Data Matrik Berukuran $i \times k$

4. Menghitung jumlah pusat *cluster* (C) untuk setiap *cluster* dengan menggunakan Persamaan 2.4 dibawah ini.

$$C_{kj} = \frac{\sum_{i=1}^n (\mu_{ki})^w X_{ij}}{\sum_{i=1}^n (\mu_{ki})^w} \quad (2.4)$$

Keterangan:

C_{jk} = Pusat *cluster* pada *cluster* ke- k dan atribut ke- j

μ_{ki} = Data partisi (matriks u) pada *cluster* ke- k dan data ke- i

X_{ij} = Data pada atribut ke- j dan data ke- i

w = Pembobot

5. Menghitung nilai obyektif (P_n) dengan Persamaan 2.5 sebagai berikut.

$$P_n = \sum_{i=1}^n \sum_{k=1}^m (\mu_{ki})^w (d_{ki})^2 \quad (2.5)$$

Keterangan:

μ_{ki} = Data partisi (matriks u) pada *cluster* ke- k dan data ke- i .

d_{ki} = Fungsi jarak Euclidean pada pusat *cluster* ke- k dan data ke- i .

w = Pembobot

P_n = Nilai obyektif pada iterasi ke- n

6. Memperbaiki derajat keanggotaan setiap data pada setiap *cluster* digunakan (matriks partisi). Proses tersebut dapat dilakukan menggunakan Persamaan 2.6 dibawah ini.

$$\mu_{ki} = \left[\sum_{j=1}^m \left(\frac{d_{ki}}{d_{ji}} \right)^{2/(w-1)} \right]^{-1} \quad (2.6)$$

Dengan,

$$d_{ik} = d(X_i - C_k) = \left[\sum_{j=1}^m (X_{ij} - C_{kj})^2 \right]^{1/2} \quad (2.7)$$

Keterangan:

μ_{ki} = Data partisi (matriks u) setiap pusat *cluster* ke- k dan data ke- i

d_{ki} = Fungsi jarak Euclidean setiap pusat *cluster* ke- k dan data ke- i

d_{ji} = Fungsi jarak Euclidean setiap pusat *cluster* ke- j dan data ke- i

w = Pembobot

X_{ij} = Data pada atribut ke- j dan data ke- i

7. Jika pusat *cluster* (C) tidak ada perubahan maka iterasi akan berhenti, selain itu perubahan nilai *error* kurang dari *threshold* yang ditentukan $|P_n - P_{n-1}| < \epsilon$. Selain dari hal tersebut diatas iterasi akan berhenti bilamana perulangan melebihi maksimum iterasi ($n > \max n$). Jika iterasi belum berhenti, maka kembali ke langkah 4.

8. Setelah iterasi berhenti maka selanjutnya ditentukan cluster tiap-tiap data berdasarkan partisi terbesar nilai matriks.

2.1.6. Penelitian yang Berkaitan

Beberapa penelitian-penelitian yang berkaitan dengan penelitian pada tulisan ini yang telah dilakukan oleh peneliti-peneliti sebelumnya dapat dilihat pada Tabel II-1 sebagai berikut.

Tabel II-1. Penelitian yang Berkaitan

No.	Nama Peneliti	Metode yang Digunakan	Hasil
1.	Ulinnuha, 2016	Pengelompokan kinerja siswa menggunakan metode <i>Fuzzy C-Means Clustering</i>	Dari hasil uji coba dengan menggunakan sistem <i>Fuzzy C-Means Clustering</i> didapatkan akurasi <i>training</i> dengan nilai RMSE sebesar 2.76×10^{-15} dan didapatkan akurasi <i>testing</i> dengan nilai RMSE sebesar 0,1165.
2	Hadi et al., 2016	Kombinasi metode Algoritma Genetika dengan <i>Fuzzy C-Means Clustering</i> dalam penentuan kinerja siswa	Hanya 61 dari 77 fitur dataset yang relevan dan dapat digunakan dalam proses clustering setelah dua algoritma ini dikombinasikan. Hasil penelitian menunjukkan bahwa algoritma genetika dan fuzzy C-means yang digabungkan memberikan hasil yang lebih baik daripada algoritma fuzzy C-means saja, dengan persentase kesesuaian rata-rata sebesar 88,89% selama pengujian yang dilakukan.
3	Waraso, 2016	Metode <i>Profile</i> digunakan sebagai Sistem pendukung keputusan evaluasi kinerja siswa.	Setiap siswa dapat melakukan penilaian melalui sistem pendukung keputusan ini, yang memungkinkan perubahan kriteria dan nilai bobot. Hasil penilaian berupa perankingan, dan siswa yang berprestasi akan diberikan penghargaan atau peringatan. Siswa yang kurang berprestasi juga akan diberikan peringatan agar mereka lebih baik melakukan tugasnya di masa depan untuk mendapatkan penilaian terbaik.

4	Yadav & Singh, 2012	Model evaluasi kinerja menggunakan teknik <i>Fuzzy C-Means Clustering</i>	Model <i>Fuzzy C-Means Clustering</i> telah memperbaiki beberapa keterbatasan dari metode tradisional yang ada, seperti metode rata-rata dan metode statistik. Algoritma <i>Fuzzy C-Means</i> adalah model terbaik untuk pemodelan kinerja dalam domain pendidikan. Oleh karena itu, algoritma <i>Fuzzy C-Means Clustering</i> berfungsi sebagai <i>benchmark</i> yang baik untuk memantau perkembangan pemodelan siswa dalam domain pendidikan.
---	---------------------	---	--

Studi ini mencoba meningkatkan kinerja atau memperbaiki algoritma pengelompokan hierarki *fuzzy C-Means* pada proses pengelompokan data kategori. Algoritma ROCK adalah algoritma pengelompokan hierarki agglomeratif untuk data kategori; algoritma pengelompokan hierarki konvensional yang menggunakan jarak antara titik dianggap kurang cocok untuk pengelompokan data kategori. Oleh karena itu, diciptakan ide baru, yaitu hubungan, untuk mengukur kesamaan antara dua titik data.

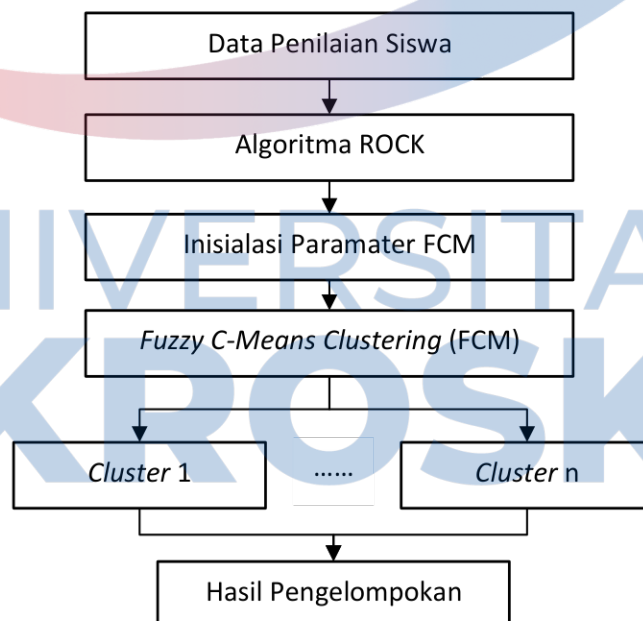
Sesuai dengan berbagai penelitian sebelumnya penelitian terdekat dengan pendekatan berbasis link dari ROCK termasuk penelitian yang terfokus pada pengembangan dan aplikasi metode clustering yang menggunakan informasi hubungan atau link antar titik data. Beberapa penelitian terkait yaitu 1. *Informative Hierarchical Clustering (IHC)* merupakan pengembangan dari pendekatan berbasis link, IHC menggabungkan informasi jaringan (network) dengan clustering hierarkis untuk menemukan struktur yang tersembunyi dalam data yang besar dan kompleks. 2. *Link-based Cluster Ensemble (LCE)* dimana Pendekatan ini menggabungkan hasil dari multiple link-based clustering untuk meningkatkan kualitas clustering, dengan mempertimbangkan informasi dari berbagai sumber link atau hubungan antar data. 3. *Spectral Clustering based on Links* dimana metode ini menggunakan pendekatan spektral untuk clustering, di mana matriks yang digunakan untuk clustering tidak hanya berdasarkan jarak

antar titik, tetapi juga mengintegrasikan informasi link atau hubungan antar titik untuk menentukan struktur cluster.

Pada penelitian ini, pengujian algoritma ROCK digunakan dalam penentuan pusat *cluster* awal pada algoritma *Fuzzy C-Means Clustering*. Algoritma ROCK akan mencari pusat *cluster* data pada dataset dengan melakukan pengukuran jarak diantara data tersebut. Hasil dari pusat *cluster* data tersebut akan dijadikan sebagai inisialisasi awal pada algoritma FCM dalam melakukan pengelompokan data dan menentukan hasil kinerja siswa.

2.2. Kerangka Konsep

Penelitian ini akan membahas tentang perbaikan algoritma *Fuzzy C-Means Clustering* dengan algoritma ROCK untuk meningkatkan akurasi pengelompokan kinerja siswa. Tahapan alur dari algoritma yang diusulkan pada penelitian ini dapat diperlihatkan pada Gambar II-7 sebagai berikut.



Gambar II-7. Tahapan Penelitian yang Diusulkan