

BAB II

KAJIAN PUSTAKA

A. Kajian Teoritis

1. Sistem

Menurut Prehanto (2020:3) menyatakan bahwa Sistem adalah kumpulan komponen yang berinteraksi satu sama lain, melalui hubungan fisik maupun non-fisik, dan bekerja bersama secara selaras untuk mencapai tujuan yang diinginkan. Sistem dapat diartikan sebagai himpunan unsur atau variabel yang terorganisir, saling berinteraksi, dan saling bergantung satu sama lain, yang dirancang untuk mendukung dan meningkatkan proses pengolahan informasi (Agustin, 2018).

Sistem adalah sekumpulan bagian yang saling berhubungan dan bekerja bersama untuk mencapai tujuan tertentu. Keberadaannya meluas di berbagai aspek kehidupan, dengan batasan dan sasaran yang terdefinisi. Pencapaian tujuan sistem melibatkan kerja sama antar komponen utamanya Khristianto (2015:6).

Berdasarkan definisi di atas, dapat disimpulkan bahwa pengertian dari Sistem adalah kumpulan komponen yang saling berinteraksi, baik fisik maupun non-fisik, bekerja secara terstruktur untuk mencapai tujuan tertentu. Sistem terdiri dari subsistem terintegrasi, memiliki batasan dan sasaran yang jelas, serta bergantung pada kerja sama antar komponen untuk menghasilkan output yang diinginkan.

2. *Data Mining*

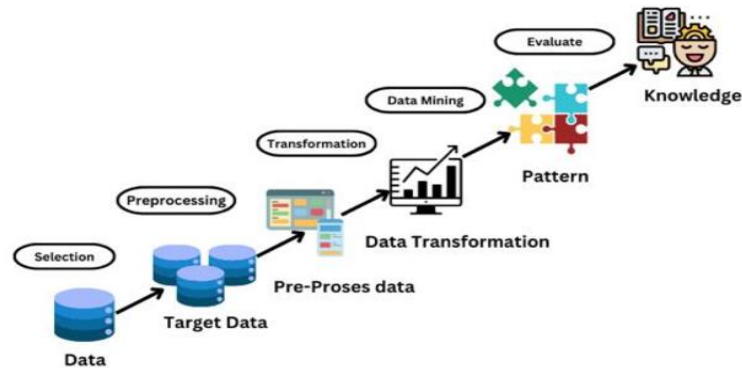
Pengertian *data mining* menurut Siregar & Puspabhuana (2021:1) menyatakan bahwa *data mining* adalah proses otomatis untuk menemukan informasi, pola, atau hubungan yang penting dan tersembunyi dari kumpulan data yang sangat besar, terutama yang tersimpan dalam *database*, proses ini juga dikenal sebagai *Knowledge Discovery in Database* (KDD) dan dapat diterapkan pada berbagai jenis data, termasuk data tabular, spasial, teks, maupun multimedia. *Data mining* merupakan teknik yang dapat membantu memprediksi fenomena suatu objek yang diteliti dengan terlebih dahulu mengetahui atau tidak mengetahui suatu kelas suatu obyek tertentu melalui masukan dari data besar (Natasuwarna, 2019).

Pengertian *data mining* menurut Arhami & Nasir (2020:1) menjelaskan *data mining* adalah proses logis dan sistematis untuk menggali pola tersembunyi, informasi bermanfaat, serta pengetahuan dari kumpulan data besar menggunakan teknik statistik, analisis matematis, dan algoritma.

Berdasarkan definisi di atas, dapat disimpulkan bahwa *data mining* adalah proses otomatis, logis, dan sistematis untuk menemukan pola tersembunyi, informasi penting, serta pengetahuan yang bermanfaat dari kumpulan data besar dengan menggunakan teknik statistik, analisis matematis, dan algoritma tertentu. Proses ini juga dikenal sebagai *Knowledge Discovery in Database* (KDD) dan dapat

diterapkan pada berbagai jenis data untuk membantu prediksi, klasifikasi, serta pengambilan keputusan yang lebih akurat.

Gambar 2. 1 Proses *Knowledge Discovery in Database*



Sumber: Sari (2024)

3. *Clustering*

Pengertian *Clustering* menurut Muhima et al., (2022:1) menyatakan bahwa *Clustering* adalah metode untuk mengelompokkan data berdasarkan karakteristik tertentu, membantu dalam analisis distribusi data. Selain itu, clustering berperan sebagai langkah awal *pre-processing* untuk metode lain seperti karakterisasi, pemilihan atribut, dan klasifikasi. *Clustering* adalah metode pembelajaran mesin yang mengelompokkan data berdasarkan kesamaan karakteristik untuk menemukan pola tersembunyi. Teknik ini memisahkan data ke dalam beberapa klaster, di mana data dalam satu klaster memiliki kemiripan, sedangkan antar klaster berbeda (Rahmadani & Nursyahira, 2025).

Pengertian *Clustering* menurut Sari et al., (2024:8) menjelaskan bahwa *Clustering* adalah tahap pengelompokan himpunan data yang belum memiliki atribut kelas yang terdefinisi. Secara konsep,

clustering bertujuan untuk memaksimalkan kesamaan dalam satu kelompok (*intra-class similarity*) dan meminimalkan kesamaan antar kelompok (*inter-class similarity*).

Berdasarkan definisi di atas, dapat disimpulkan bahwa *Clustering* adalah metode pembelajaran mesin yang digunakan untuk mengelompokkan data berdasarkan karakteristik tertentu. Teknik ini membantu dalam analisis distribusi data, menemukan pola tersembunyi, serta berfungsi sebagai tahap awal dalam berbagai metode analisis lainnya seperti karakterisasi, pemilihan atribut, dan klasifikasi. *Clustering* juga bertujuan untuk memaksimalkan kesamaan dalam satu kluster dan meminimalkan kesamaan antar kluster, sehingga data lebih terstruktur dan mudah dianalisis.

4. K-Means

Pengertian K-Means menurut Rizka et al., (2024:10) menyatakan bahwa K-Means adalah salah satu teknik *clustering* pada data *mining*, proses pemodelan tanpa supervisi, dan metode pengelompokan data secara partisi. K-Means adalah metode *clustering* berbasis jarak yang sering digunakan dalam data *mining* untuk mengelompokkan data berdasarkan kesamaan karakteristik. Data yang mirip dikelompokkan dalam satu kluster, sementara data yang berbeda dimasukkan ke kluster lain (Rahmadani & Nursyahira, 2025).

Pengertian K-Means menurut Prianto & Bunyamin (2020:15) menjelaskan bahwa K-Means *clustering* dapat didefinisikan sebagai proses pengelompokan objek ke dalam kelompok-kelompok yang memiliki kemiripan dalam beberapa aspek. Berdasarkan definisi di atas, dapat disimpulkan bahwa K-Means adalah metode *clustering* berbasis jarak yang digunakan dalam data *mining* untuk mengelompokkan data secara partisi tanpa supervisi. Metode ini membagi sekumpulan data menjadi K klaster yang berbeda dan tidak tumpang-tindih, di mana data dengan karakteristik serupa dikelompokkan dalam satu klaster, sementara data yang berbeda dimasukkan ke klaster lain.

Gambar 2. 2 Perhitungan Jarak K-Means

$$d_{xy} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

Sumber: Marisa et al., (2021)

5. Bahasa Pemrograman Python

Pengertian Python menurut Enterprise (2019:1) menyatakan bahwa Python adalah bahasa pemrograman yang mudah dipelajari, memiliki sintaks yang jelas, serta berfokus pada keterbacaan kode, sehingga mempermudah pemahaman dan pengembangannya. Python adalah bahasa pemrograman interpretatif multiguna yang berfokus pada keterbacaan kode, memiliki sintaks yang jelas, serta didukung oleh pustaka standar yang luas dan komunitas besar. Python mendukung berbagai paradigma pemrograman, termasuk berorientasi objek, imperatif, dan fungsional (Syahrudin & Kurniawan, 2018).

Menurut Maesaroh et al., (2024:1) Python adalah bahasa pemrograman tingkat tinggi yang dikembangkan pada 1991 oleh Guido van Rossum, dengan desain yang menekankan keterbacaan kode dan sintaks sederhana, menjadikannya populer bagi pemula dan pengembang berpengalaman.

Berdasarkan definisi di atas, dapat disimpulkan bahwa Python adalah bahasa pemrograman tingkat tinggi yang dikembangkan oleh Guido van Rossum pada tahun 1991. Python memiliki sintaks yang sederhana dan jelas, serta berfokus pada keterbacaan kode, sehingga mudah dipelajari dan digunakan oleh pemula maupun pengembang berpengalaman. Sebagai bahasa pemrograman interpretatif multiguna, Python mendukung berbagai paradigma pemrograman seperti berorientasi objek, imperatif, dan fungsional.

a. Flask *Framework*

Pengertian Flask menurut Maesaroh et al., (2024:145) menyatakan bahwa Flask adalah *micro-framework* Python yang dirancang untuk menyederhanakan pengembangan aplikasi web. Flask adalah *micro-framework* Python yang ringan, fleksibel, dan dapat diperluas dengan ekstensi sesuai kebutuhan. Mengandalkan Werkzeug, Jinja2, dan Click, Flask tidak memiliki fitur bawaan seperti *database* atau autentikasi, memberi kebebasan penuh kepada pengembang dalam memilih teknologi yang digunakan (Darmawan, 2022).

Menurut Stouffer et al., (2018:6) Flask adalah *microframework* dengan API yang dirancang baik, menyediakan fungsionalitas minimal untuk pengembangan web tanpa fitur bawaan seperti *database*, *cache*, atau keamanan.

Berdasarkan definisi di atas, dapat disimpulkan bahwa Flask adalah *micro-framework* Python yang ringan dan fleksibel, dirancang untuk menyederhanakan pengembangan aplikasi web. Framework ini menyediakan fungsionalitas inti tanpa fitur bawaan seperti *database*, *cache*, atau autentikasi, tetapi dapat diperluas dengan berbagai ekstensi sesuai kebutuhan pengembang. Dengan mengandalkan Werkzeug, Jinja2, dan Click, Flask memberikan kebebasan penuh dalam memilih teknologi yang digunakan dalam pengembangan web.

b. MySQL

Menurut Surya & Aminuddin (2024:38) Menyatakan Bahwa MySQL adalah sistem manajemen basis data (DBMS) *open source* yang menggunakan bahasa SQL. MySQL mendukung *multi-user* dan *multi-thread*, tersedia di berbagai platform seperti Windows, Linux, dan macOS. Dikembangkan oleh perusahaan Swedia MySQL AB, MySQL tersedia dalam versi gratis (GPL) dan versi komersial. MySQL adalah database server populer yang termasuk RDBMS (*Relational Database Management System*). MySQL menggunakan bahasa SQL yang telah distandarkan oleh

ANSI dan mendukung integrasi dengan bahasa pemrograman seperti PHP (Hermiati et al., 2021).

Menurut Namruddin et al., (2023:2) Menyatakan bahwa MySQL merupakan sistem manajemen basis data (DBMS) *open source* yang tersedia dalam dua jenis lisensi, yaitu perangkat lunak bebas (*Free Software*) dan perangkat lunak berpemilik dengan penggunaan terbatas (*Shareware*). Dengan lisensi GNU *General Public License* (GPL), MySQL dapat digunakan secara gratis baik untuk keperluan pribadi maupun komersial tanpa perlu membayar biaya lisensi.

Berdasarkan definisi di atas, dapat disimpulkan bahwa MySQL adalah DBMS *open source* berjenis RDBMS yang menggunakan bahasa SQL standar ANSI. MySQL mendukung *multi-user, multi-thread*, dan dapat digunakan di berbagai platform. Tersedia dalam versi gratis (GPL) dan komersial, serta mendukung integrasi dengan bahasa pemrograman seperti PHP.

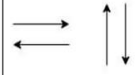







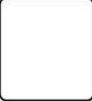
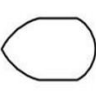
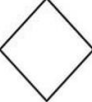

6. *Flowchart*

Menurut Sitorus (2015:14) menyatakan bahwa *Flowchart* adalah diagram alir yang digunakan untuk menggambarkan urutan logika dalam suatu prosedur pemecahan masalah. *Flowchart* merupakan gambar atau bagan yang memperlihatkan urutan atau langkah-langkah dari suatu program dan hubungan antar proses beserta pernyataannya (Fauzi, 2020).

Menurut Suyanto (2018:259) *Flowchart* adalah representasi visual dari aliran data dalam pemrosesan informasi, operasi dalam sistem, dan urutan proses yang dilakukan.

Berdasarkan definisi di atas, dapat disimpulkan bahwa *flowchart* adalah diagram alir yang menggambarkan urutan logika dan langkah-langkah dalam suatu prosedur pemecahan masalah. *Flowchart* digunakan untuk menunjukkan aliran data, operasi dalam sistem, serta hubungan antar proses dalam suatu program, yang disajikan secara visual untuk memudahkan pemahaman dan komunikasi.

Tabel 2. 1 Simbol *Flowchart*

	Flow Simbol yang digunakan untuk menggabungkan antara simbol yang satu dengan simbol yang lain. Simbol ini disebut juga dengan Connecting Line.		Input/output Simbol yang menyatakan proses input atau output tanpa tergantung peralatan.
	On-Page Reference Simbol untuk keluar - masuk atau penyambungan proses dalam lembar kerja yang sama.		Manual Operation Simbol yang menyatakan suatu proses yang tidak dilakukan oleh komputer.
	Off-Page Reference Simbol untuk keluar - masuk atau penyambungan proses dalam lembar kerja yang berbeda.		Document Simbol yang menyatakan bahwa input berasal dari dokumen dalam bentuk fisik, atau output yang perlu dicetak.
	Terminator Simbol yang menyatakan awal atau akhir suatu program.		Predefine Proses Simbol untuk pelaksanaan suatu bagian (sub-program) atau prosedur.
	Process Simbol yang menyatakan suatu proses yang dilakukan komputer.		Display Simbol yang menyatakan peralatan output yang digunakan.
	Decision Simbol yang menunjukan kondisi tertentu yang akan menghasilkan dua kemungkinan jawaban, yaitu ya dan tidak.		Preparation Simbol yang menyatakan penyediaan tempat penyimpanan suatu pengolahan untuk memberikan nilai awal.

Sumber: Fauzi (2020)

7. *Data Flow Diagram (DFD)*

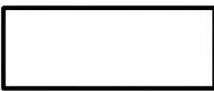
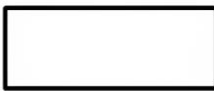






Menurut Rukmana et al., (2023:112) menyatakan bahwa DFD (*Data Flow Diagram*) adalah representasi grafis yang menggambarkan

aliran dan transformasi informasi dalam suatu sistem, dari masukan (*input*) hingga keluaran (*output*). Diagram ini sangat berguna dalam pemodelan fungsional dan lebih sesuai untuk sistem berbasis program terstruktur, di mana proses dapat dibagi menjadi fungsi dan prosedur yang lebih kecil. *Data Flow Diagram* (DFD) adalah diagram yang menggambarkan aliran data dalam sistem, termasuk proses *input*, pemrosesan, dan *output*. DFD membantu memvisualisasikan hubungan antar subsistem serta memudahkan komunikasi sistem yang dirancang kepada pengguna (Soufitri, 2019).

Menurut Wijayanto et al., (2024:70) *Data Flow Diagram* (DFD) adalah representasi sistem yang menggambarkan aliran data dari satu proses ke proses lainnya serta lokasi penyimpanannya, membantu dalam memahami struktur dan hubungan antar komponen dalam sistem.

Berdasarkan definisi di atas, dapat disimpulkan bahwa *Data Flow Diagram* (DFD) adalah representasi grafis yang menunjukkan aliran data dalam suatu sistem, dari *input* hingga *output*. DFD membantu memahami struktur sistem, hubungan antar komponen, dan mempermudah komunikasi dalam perancangan sistem.

Tabel 2. 2 DFD Diagram

Keterangan	DeMarco and Yourdan Simbol	Gane and Sarson Simbol
Entitas Luar		
Proses		
Aliran data (data flow)		
Simpan data		

Sumber: Soufitri (2019)

8. *Entity Relationship Diagram (ERD)*


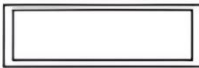
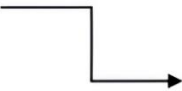






Menurut Kusuma et al., (2024:135) menyatakan bahwa *Entity Relationship Diagram (ERD)* digunakan untuk merepresentasikan aspek-aspek data, termasuk entitas, atribut, dan hubungan (*relationship*) di dalam suatu sistem. *Entity Relationship Diagram (ERD)* adalah diagram yang digunakan untuk merancang dan menggambarkan hubungan antar tabel dalam basis data (Apriliah et al., 2019).

Menurut Romadloni & Miswanto (2024:14) *ERD (Entity Relationship Diagram)* adalah alat visual yang merepresentasikan struktur basis data dalam bentuk abstrak, terdiri dari entitas serta hubungan antar entitas, sehingga memudahkan pemahaman sistem penyimpanan data.

Berdasarkan pernyataan di atas, dapat disimpulkan bahwa *Entity Relationship Diagram (ERD)* adalah alat visual yang digunakan

untuk merancang dan menggambarkan hubungan antar tabel dalam basis data. ERD merepresentasikan entitas, atribut, dan hubungan antar entitas dalam suatu sistem, memudahkan pemahaman terhadap struktur dan sistem penyimpanan data.

Tabel 2. 3 ERD Diagram

No	Simbol/Notasi	Arti
1		Entity Type (Strong Entity)
2		Weak Entity
3		Relationship
4		Identifying relationship
5		Attribute (Simple)
6		Key Attribute
7		Multivaluated Attribute
8		Composite Attribute
9		Derived Attribute

Sumber: Apriliah et al., (2019)

B. Kajian Empiris

Penelitian yang membahas tentang sistem pendukung keputusan dalam pemilihan jurusan telah banyak diterapkan sehingga dijadikan bahan referensi. Perancangan dan pengembangan sistem semacam ini memerlukan penelitian yang tepat dari berbagai sumber penelitian yang relevan.

Penelitian sebelumnya oleh Nengsih et al., (2021) membahas tentang analisis pengelompokan penentuan jurusan siswa SMA menggunakan algoritma K-Means. Penelitian ini dilakukan di SMA Negeri 21 Bandung dengan menggunakan data nilai rapor semester 1 sampai 5. Hasil pengujian menunjukkan bahwa algoritma K-Means mampu membentuk tiga *cluster*, yaitu siswa yang tidak diterima di dua jurusan, siswa yang masuk jurusan IPS, dan siswa yang masuk jurusan IPA. Proses *clustering* menggunakan RapidMiner menunjukkan efisiensi waktu dan hasil yang lebih akurat dibanding metode manual yang dilakukan oleh pihak sekolah.

Penelitian selanjutnya dilaksanakan oleh Anggraini & Sau'da (2023) mengembangkan sistem rekomendasi pemilihan konsentrasi di Program Studi Informatika Universitas Bina Darma menggunakan algoritma K-Means. Data yang digunakan adalah data akademik mahasiswa dari semester 1 sampai 4. Hasil *clustering* memberikan rekomendasi konsentrasi yang sesuai dengan kemampuan mahasiswa dan membantu pihak prodi dalam memberi arahan akademik. Penelitian ini menekankan

pentingnya pengelompokan berbasis performa akademik dalam sistem pendukung keputusan pendidikan.

Penelitian oleh Ariyanto et al., (2024) mengembangkan sistem rekomendasi untuk pengelompokan kelas siswa baru di SMAN 1 Durenan menggunakan metode K-Means. Sistem ini memproses data nilai rapor, nilai ujian, IQ, gaya belajar, dan nilai akademik lainnya. Hasil *clustering* memberikan referensi yang optimal dalam pembagian kelas yang efektif dan efisien. Penelitian ini membuktikan bahwa K-Means dapat digunakan dalam sistem pendukung keputusan pada konteks pembelajaran di tingkat SMA.

Penelitian yang dilakukan oleh Palevi & Indra (2024) menerapkan algoritma K-Means dengan pendekatan *Active Learning* untuk membantu siswa SMA menentukan jurusan kuliah. Data yang digunakan adalah gabungan antara hasil kuisisioner minat dan bakat serta nilai rapor siswa. Hasil *clustering* menghasilkan 12 klaster yang mewakili berbagai rumpun ilmu. Penelitian ini menunjukkan bahwa pendekatan *Active Learning* dapat meningkatkan akurasi hasil klasifikasi bidang studi yang sesuai dengan minat siswa.

Penelitian yang dilaksanakan oleh Afifuddin & Nurjanah (2019) menggabungkan algoritma K-Means dan Apriori untuk merekomendasikan mata kuliah peminatan di jurusan Teknik Informatika Universitas Telkom. Algoritma K-Means digunakan untuk mengelompokkan mahasiswa ke dalam kelompok keahlian, sedangkan Apriori digunakan untuk membentuk aturan rekomendasi pemilihan mata

kuliah. Hasilnya menunjukkan bahwa penggabungan kedua metode tersebut mampu memberikan rekomendasi pemilihan mata kuliah secara akurat dan efisien.

Berdasarkan penelitian-penelitian di atas, dapat disimpulkan bahwa sistem pendukung keputusan dengan metode K-Means *Clustering* dapat diterapkan secara luas dalam konteks pendidikan, baik untuk pemilihan jurusan maupun pengelompokan siswa. Pendekatan ini terbukti efektif dalam menghasilkan pengelompokan data yang bermanfaat untuk pengambilan keputusan akademik.

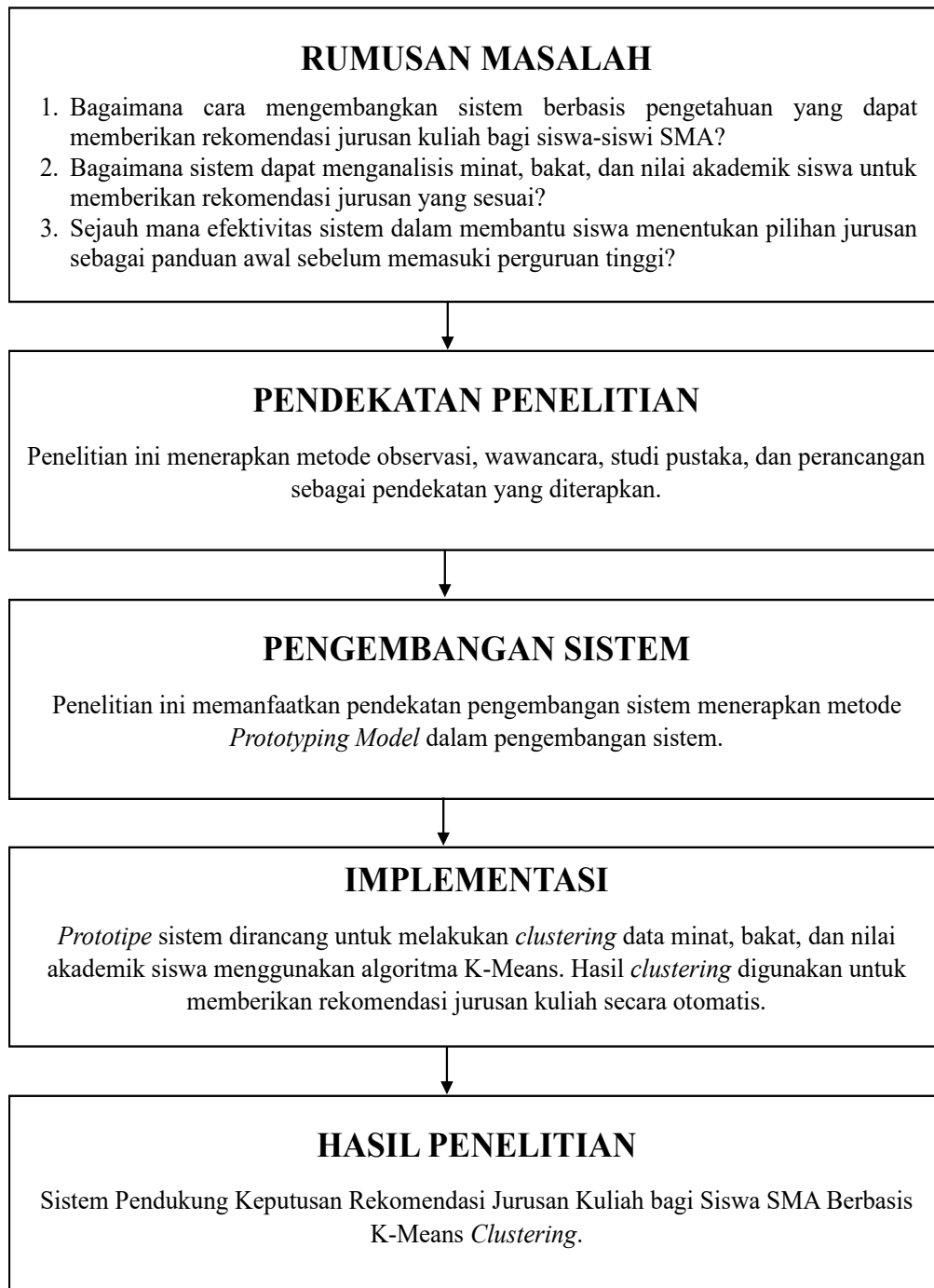
C. Kerangka Berpikir

Permasalahan utama yang diangkat dalam penelitian ini adalah kesulitan yang dihadapi siswa-siswi SMA dalam menentukan jurusan kuliah yang sesuai dengan minat, bakat, dan kemampuan akademik mereka. Ketidaktepatan dalam pemilihan jurusan seringkali berujung pada rendahnya motivasi belajar, kesulitan mengikuti perkuliahan, bahkan keinginan untuk berpindah jurusan, yang pada akhirnya dapat menghambat pencapaian prestasi dan tujuan karier mereka.

Penelitian ini mengembangkan model SPK yang memanfaatkan algoritma K-Means untuk mengelompokkan siswa berdasarkan kesamaan data minat, bakat, dan nilai akademik mereka. Model ini akan memproses data yang diperoleh melalui pengisian kuesioner dan dokumen akademik untuk mengidentifikasi pola dan menghasilkan cluster siswa yang memiliki karakteristik serupa. Setiap cluster akan diasosiasikan dengan jurusan kuliah yang paling sesuai berdasarkan pola data yang ditemukan.

Objek dalam penelitian ini adalah siswa-siswi kelas XII SMA yang sedang berada dalam tahap penentuan jurusan kuliah. Luaran dari penelitian ini adalah sebuah *prototype* sistem pendukung keputusan berbasis web yang mampu memberikan rekomendasi jurusan kuliah secara otomatis berdasarkan hasil *clustering*. Sistem ini diharapkan dapat menjadi alat bantu yang berguna bagi siswa dalam mengambil keputusan akademik.

Berikut adalah bagan kerangka berpikir pada penelitian ini.



Gambar 2.1 Kerangka Berpikir