

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi informasi yang begitu pesat pada zaman Industri 4.0, terutama perkembangan pada teknologi *Data Science*. Peran *data science* pada industri 4.0 di tahun 2021 sangat dibutuhkan dan memberikan dampak yang begitu kuat dalam ekosistem dunia. Industri 4.0 merupakan sebuah revolusi dari inovasi di bidang Elektronik dan Teknologi Informasi. Konsep Industri 4.0 adalah realitas baru dari ekonomi modern, karena inovasi dan pengembangan teknologi memainkan peran penting dalam setiap organisasi (Ślusarczyk, 2018). Terdapat beberapa macam teknologi yang akan menjadi pilar utama dalam revolusi Industri 4.0, yaitu *Big Data*, *Internet of Things*, *Cloud Computing*, *Addictive Manufacturing*, dan salah satunya *Artificial Intelligence*.

*Artificial Intelligence* (AI) adalah salah satu bagian ilmu komputer yang mempelajari bagaimana membuat mesin (komputer) dapat melakukan pekerjaan seperti dan sebaik yang dilakukan oleh manusia bahkan bisa lebih baik daripada yang dilakukan manusia (Dahria, 2008). *Artificial Intelligence* dalam ilmu teknologi terbagi dalam 7 cabang, yaitu *Machine Learning*, *Natural Language Processing* (NLP), *Expert System*, *Vision*, *Speech*, *Planning*, dan *Robotic*.

*Machine learning* merupakan cabang dari *Artificial Intelligence*. *Machine Learning* adalah disiplin ilmu yang mencakup perancangan dan pengembangan algoritma yang memungkinkan komputer untuk mengembangkan perilaku yang didasarkan pada data empiris (I. Kambey *et al.*, 2020). *Machine machine learning* terbagi menjadi tiga kategori: Supervised Learning, Unsupervised Learning, Reinforcement Learning (Somvanshi & Chavan, 2016).

Teknik yang digunakan oleh Supervised Learning adalah metode klasifikasi di mana kumpulan data sepenuhnya diberikan label untuk mengklasifikasikan kelas yang tidak dikenal. Sedangkan teknik Unsupervised Learning sering disebut cluster dikarenakan tidak ada kebutuhan untuk pemberian label dalam kumpulan data dan hasilnya tidak mengidentifikasi contoh di kelas yang telah ditentukan (Thupae, Isong, Gasela, & Abu-Mahfouz, 2018). Sedangkan Reinforcement Learning biasanya berada di antara Supervised Learning dan Unsupervised Learning (Board, 2017), teknik ini bekerja dalam lingkungan yang dinamis di mana konsepnya harus menyelesaikan tujuan tanpa adanya pemberitahuan dari komputer secara eksplisit jika tujuan tersebut telah tercapai (Das & Nene, 2017).

Analisis *cluster* adalah analisis untuk mengelompokkan elemen yang mirip sebagai objek penelitian untuk menjadi kelompok (*cluster*) yang berbeda dan *mutually exclusive*. Analisis *cluster* termasuk dalam analisis statistic *multivariat* metode interdependen, dan oleh karena itu tujuan analisis *cluster* tidak untuk menghubungkan ataupun membedakan dengan sampel/variabel lain. Analisis *cluster* berguna untuk

meringkas data dengan jalan mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu diantara objek-objek yang akan diteliti (Robinson Sitepu, Irmeilyana, dan Berry Gultom, 2011).

Penentuan jurusan akan berdampak terhadap kegiatan akademik selanjutnya dan mempengaruhi pemilihan bidang ilmu atau studi bagi siswa-siswi yang ingin melanjutkan ke perguruan tinggi nantinya. Penentuan jurusan yang dilakukan selama ini mempunyai banyak kelemahan, antara lain berdasarkan keinginan siswa tanpa melihat latar belakang nilai akademisnya. Sehingga jurusan yang dipilih terkadang menjadi masalah bagi siswa di kemudian hari, sebagai contoh nilai akademik yang tidak maksimal, pemilihan program studi saat melanjutkan ke jenjang perguruan tinggi yang terkendala akibat jurusan SMA yang tidak sesuai, dan lain-lain.

Dalam penelitian ini akan menganalisis kasus pengelompokkan *K-Means Cluster* Siswa Kelas 12 yang berjurusan IPA/IPS di SMA N 2 Palembang berdasarkan Minat Dan Bakat siswa/siswi untuk dijadikan rekomendasi untuk memilih jurusan ke jenjang berikutnya. *K-Means Cluster* adalah salah satu metode dan *clustering non-hierarki* yang berusaha mengelompokkan data ke dalam suatu *cluster* sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu *cluster* yang sama. Ukuran kesamaan yang digunakan adalah ukuran jarak antar objek. Kedua objek yang memiliki jarak paling dekat akan bergabung menjadi satu cluster. Kedekatan jarak yang dimiliki menunjukkan bahwa kedua objek tersebut memiliki tingkat kesamaan karakteristik (Sibarani, 2018). Berdasarkan latar belakang yang telah dijelaskan, dapat

dirumuskan masalah yang akan dibahas pada penelitian ini, yaitu melakukan klusterisasi minat dan bakat di SMA N 2 Palembang menggunakan metode *K-Means Clustering*. Ada 24 kriteria yang bisa diteliti yaitu (minat, jurusan, hobi, sifat, bakat, pendidikan agama, pendidikan pancasila dan kewarganegaraan, bahasa Indonesia, matematika, sejarah Indonesia, bahasa Inggris, seni budaya, pendidikan jasmani olahraga dan kesehatan, fisika IPA, matematika peminatan IPA, kimia IPA, biologi IPA, ekonomi IPA, bahasa Inggris IPA, ekonomi IPS, sosiologi IPS, sejarah IPS, geografi IPS, bahasa Inggris IPS). Diharapkan tingkat persentase kelulusan dalam menentukan jurusan Perguruan Tinggi Negeri pada siswa-siswi SMA Negeri 2 Palembang lebih tinggi.

### **1.2 Rumusan Masalah**

Berdasarkan permasalahan yang telah dibahas dapat diuraikan rumusan masalah dalam penelitian ini yaitu bagaimana penerapan metode *K-means Clustering* dalam mengelompokkan minat dan bakat siswa SMA N 2 Palembang untuk merekomendasikan jurusan ke perguruan tinggi.

### **1.3 Batasan Masalah**

Agar mendekati tujuan yang diharapkan, maka diperlukan pembatasan masalah sebagai berikut.

1. Dibatasi pada analisis *cluster K-Means* beserta teori-teori yang mendukung.

2. Studi kasus penelitian ini dibatasi pada data kuisioner minat dan bakat siswa kelas 12 di SMA N 2 Palembang tahun 2021 - 2022 berdasarkan Minat dan bakat serta nilai rapor siswa.
3. Analisis data dilakukan dengan data yang diperoleh dari kuisioner yang telah di bagikan melalui *google.form* dan dibantu melalu wakil kurikulum sekolah.
4. Penelitian ini menggunakan *RapidMiner* sebagai *tools data mining* untuk mendukung *clustering* menggunakan Metode *K-means*.

#### **1.4 Tujuan Penelitian**

Tujuan yang ingin dicapai melalui penelitian ini adalah sebagai berikut.

1. Membuat klasterisasi minat dan bakat siswa-siswi SMA Negeri 2 Palembang
2. Membantu siswa-siswi SMA Negeri 2 Palembang agar tidak ragu memilih jurusan
3. Mengetahui variabel yang memberikan perbedaan paling besar pada *cluster* yang terbentuk dan jurusan mana yang memiliki jumlah terbanyak dalam variabel tersebut.

#### **1.5 Manfaat Penelitian**

Pada Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

1. Membantu pihak SMA Negeri 2 Palembang mengarahkan penjurusan kuliah sesuai minat dan bakat yang dimiliki siswa-siswinya.
2. Membantu siswa-siswi SMA Negeri 2 Palembang untuk memilih jurusan kuliah.

## **1.6 Metodologi Penelitian**

### **1.6.1. Waktu dan Tempat Penelitian**

Tempat penelitian dilaksanakan di SMA Negeri 2 Palembang yang berlokasi di Jl. Puncak Sekuning No.84, Lorok Pakjo, Kec. Ilir Bar. I, Kota Palembang, Sumatera Selatan 30137, Dan waktu penelitian ini dilakukan dimulai dari September – Desember 2021.

### **1.6.2. Alat dan Bahan**

Adapun alat dan bahan yang penulis gunakan dalam penelitian ini yaitu *hardware* (perangkat keras) dan *software* (perangkat lunak).

#### **a. Hardware**

1. Laptop
2. Printer
3. *HardDisk Drive* 500gb

#### **b. Software**

1. *Operating System Windows 10*

2. Microsoft Word
3. Microsoft Excel
4. Web Browser Microsoft Edge
5. *RapidMiner Studio*

### **1.6.3. Metode Penelitian**

Metode penelitian yang digunakan pada penelitian ini ada 2 yaitu *Dataset* dan metode analisa data yang merupakan bagian terpenting dalam pemrosesan data. Berikut penjelasan dari metode penelitian yang akan digunakan :

#### **1. *Dataset***

Berdasarkan masalah itu muncul lah ide penulis untuk menganalisis penelitian ini menggunakan data nilai rapor siswa kelas XII dari semester satu sampai empat. Dari data tersebut dapat di kalkulasikan di Microsoft Excel dan dituangkan ke dalam aplikasi Rapid Miner dengan menggunakan metode Clustering dan algoritma K-Means. Dari hasil tersebut peneliti akan mendapatkan hasil berupa rekomendasi bidang studi perguruan tinggi. Hasil ini bertujuan untuk membantu siswa dalam menentukan bidang studi apa yang sesuai dengan nilai raport yang mereka miliki.

#### **2. *Metode Analisis Data***

Dalam penelitian ini metode yang digunakan untuk menganalisa data dalam penerapan data mining ini menggunakan proses tahapan *Knowledge*

*Discovery in Databases* (KDD). Data mining, sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso, 2007). Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Fayyad, 1996):

a. *Data Selection*

Pemilihan data dari sekumpulan data operasional yang perlu dilakukan sebelum penggalian informasi dalam KDD dimulai. Data dari hasil seleksi digunakan untuk proses data mining, disimpan pada suatu berkas, terpisah dari basis data operasional.

b. *Pre-processing/Cleaning*

Tahap cleaning ini membersihkan data yang duplikat, mengisi data yang kurang, data inkonsisten akan dilakukan pemeriksaan, kemudian dilakukan pemeriksaan pada data, seperti kesalahan dalam menulis (tipografi). Pada tahap ini juga dilakukan proses *enrichment*, yaitu suatu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan, seperti data atau informasi eksternal.

c. *Transformation*

Tahap ini dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

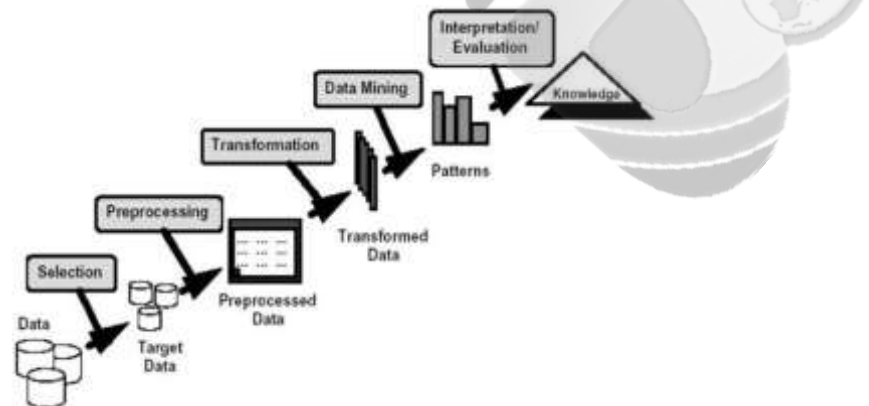


d. *Data Mining*

Merupakan proses mencari pola atau informasi yang sesuai dalam data yang terpilih dengan menggunakan teknik atau metode tertentu. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD.

e. *Interpretation/Evaluation*

Pola hasil informasi dari keluaran dalam proses data mining perlu dirubah dalam bentuk yang mudah dimengerti oleh orang yang bersangkutan. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

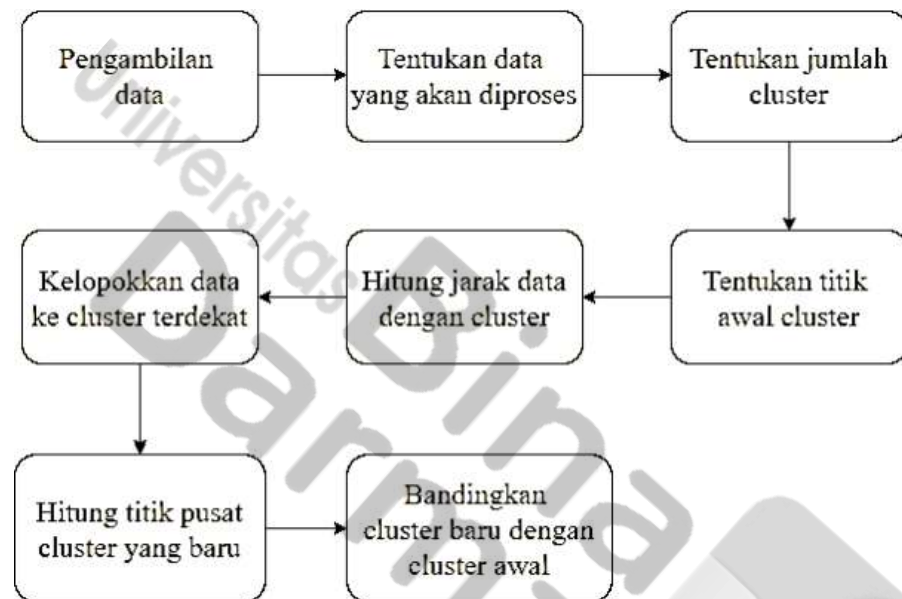


**Gambar 1.1** Proses *Knowledge Discovery in Database* (KDD)

### 3. Penerapan Metode

Penerapan metode K-means pada penelitian *Clustering* data rekomendasi penentuan jurusan ke perguruan tinggi menggunakan metode K-

*Means* studi kasus SMA Negeri 2 Palembang dilakukan dalam beberapa tahap seperti pada gambar dibawah ini .



**Gambar 1.2** Penerapan Metode *K-Means*

Berikut penjelasan tiap tahapan penerapan metode *K-Means* yang dilakukan pada penelitian ini.

#### 1. Pengambilan Data

Data penelitian ini diperoleh dari Sekolah. Data yang diperoleh disesuaikan dengan spesifikasi basis data. Data tersebut kemudian disimpan ke dalam pangkalan data yang telah dibuat.

#### 2. Seleksi Dataset

Menentukan data yang ingin diproses berdasarkan minat dan bakat serta nilai rapor semester 1-4. Dataset yang digunakan

bersifat kumulatif, sehingga data yang dipilih adalah data siswa kelas 12 yang diberikan oleh SMA Negeri 2 Palembang.

### 3. Menentukan Jumlah Klaster

Menentukan berapa jumlah klaster yang ingin dibentuk pada tiap proses. Pada penelitian dibatasi jumlah klaster yaitu 5 klaster.

### 4. Menentukan Titik Awal Klaster (Centroid)

Titik pusat klaster awal, atau dapat juga disebut centroid awal, ditentukan secara random berdasarkan jumlah klaster dan jumlah data yang akan diproses.

### 5. Menghitung Jarak Setiap Data dengan Titik Pusat Klaster

Jarak antar setiap data dengan masing-masing klaster dihitung menggunakan rumus Euclidean Distance (D) seperti disajikan pada Persamaan berikut.

$$D_{(i,j)} = \sqrt{\sum_{k=1}^n (X_{ik} - C_{jk})^2} \quad \dots \dots \dots (2)$$

yang mana

D = jarak klaster

X<sub>ik</sub> = nilai data (i,k)

C<sub>jk</sub> = nilai centroid (j,k)

n = jumlah klaster

### 6. Pengelompokan Data Berdasarkan Klaster Terdekat

Perhatikan kluster mana yang memiliki jarak terdekat dengan data. kemudian kelompokkan data ke dalam kluster tersebut.

#### 7. Menghitung Titik Pusat Kluster yang Baru

Setelah semua data dikelompokkan ke dalam kluster, hitung titik pusat kluster yang baru dengan menghitung rata-rata jarak data dengan titik pusat kluster menggunakan Persamaan berikut.

$$C_i = \frac{\sum d_i}{n_k} \dots \dots \dots (3)$$

#### 8. Bandingkan Kluster Baru dengan Kluster Awal

Jika kluster baru yang dibentuk memiliki centroid yang berbeda dengan kluster awal, ulangi kembali proses mulai dari tahap 5. Jika centroid kluster baru sudah sama dengan kluster sebelumnya, proses dapat dihentikan dan diperoleh hasil akhir clustering.

### 4. Metode Pengumpulan Data

#### a. Primer

##### 1. Observasi

Penulis melakukan pengamatan langsung ke lapangan untuk mendapatkan data-data yang jelas tentang penelitian ini.

##### 2. Wawancara

Penulis melakukan wawancara dengan pihak sekolah SMA Negeri 2 Palembang yang berhubungan dengan topik penelitian yang diangkat.

### 3. Kuisioner

Peneliti melakukan pengumpulan data yang dilakukan dengan cara melakukan mengajukan pertanyaan tertutup yang harus dijawab oleh siswa-siswi yang ada di SMA Negeri 2 Palembang berkaitan dengan minat dan bakat mereka serta nilai rata-rata mata pelajaran yang nantinya akan dikelompokkan oleh peneliti di dalam penelitian ini.

### b. Sekunder

Literatur atau studi pustaka Penulis mengumpulkan data dan informasi terkait dengan penelitian berupa jurnal dan sumber bacaan lain yang dijadikan sebagai bahan acuan atau pedoman.

## 1.7 Sistematika Penulisan

Penulisan Sistematis skripsi ini diartikan sebagai pedoman ataupun garis besar penulisan laporan penelitian yang dilakukan serta secara jelas mendeskripsikan isi laporan penelitian guna menjalin hubungan antara bab pertama sampai akhir. Sistem penulisan laporan penelitian ini terdiri atas:

## **BAB I        PENDAHULUAN**

Bab ini membahas mengenai latar belakang, rumusan masalah, batasan masalah, tujuan serta kaidah penelitian, metode penelitian yang

digunakan, metode pengumpulan data, metode analisis data, metode pengujian serta sistematika penulisan.

## **BAB II      TINJAUAN PUSTAKA**

Bab kedua membahas tentang landasan teori, yaitu teori–teori umum serta khusus guna membantu penulisan skripsi ini. Masalah yang meliputi di dalamnya merupakan pembahasan mengenai apa itu *Data Mining, Clustering, RapidMiner, K-Means* beserta teori yang menjadi acuan dari penelitian sebelumnya.

## **BAB III     ANALISIS DATA MINING**

Bab ketiga ini mengulas mengenai analisis yang dilakukan dalam Proses *data mining*, untuk memberikan gambaran mengenai masalah yang dihadapi beserta solusinya.

## **BAB IV     HASIL DAN PEMBAHASAN**

Bab ini membahas dan menguraikan hasil dari proses analisis *Data mining* yang telah dibuat beserta pembahasannya.

## **BAB V      KESIMPULAN DAN SARAN**

Pada Bab ini berisi Kesimpulan beserta saran yang bermanfaat bagi semua pihak.