

# 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* di Industri 4.0 pada tahun 2021 sangat dibutuhkan dan berdampak sangat kuat terhadap ekosistem global. Industri 4.0 merupakan revolusi inovasi di bidang elektronika dan teknologi informasi. Konsep Industri 4.0 merupakan realitas baru ekonomi modern, karena inovasi dan pengembangan teknologi memainkan peran penting di semua 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 learning* memiliki 3 macam teknik pada umumnya, yaitu *Supervised Learning* yang mempelajari pada data yang ada 2 dan mencari pola berdasarkan *dataset* yang akan dijadikan sebagai acuan untuk data-data selanjutnya, *Reinforcement Learning* yang menempatkan *agent* ke sebuah *environment* yang tidak diketahui tempatnya dan akan mengeksplorasi keseluruhan sampai menemukan istilah *reward* dan *error*, dan *Unsupervised Learning* bersifat deskriptif dan berguna untuk mengkategorikan atau mengelompokkan data (Rauhan, 2019).

Analisis *Cluster* merupakan teknik yang bervariasi yang tujuan utamanya adalah untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *Cluster* mengklasifikasikan objek sehingga objek yang paling mirip dengan objek lain berada dalam *cluster* yang sama (Ediyanto et al., 2013). Ciri-ciri pengelompokan yang baik adalah mempunyai kesamaan (homogenitas) yang tinggi antar anggota dalam satu *cluster* (*within cluster*), dan mempunyai ketaksamaan (heterogenitas) yang tinggi antar *cluster* (*between cluster*) (Yulianto & Hidayatullah, 2014).

*Coronavirus (COVID-19)* adalah jenis baru virus yang diidentifikasi pertama kali di Wuhan, China pada bulan Desember 2019 (Li et al., 2020). Infeksi virus ini kemudian menyebar ke seluruh China dan negara-negara lain di seluruh dunia yang ditetapkan sebagai pandemi oleh Organisasi Kesehatan Dunia

(Swastika, 2020) . Saat ini seluruh dunia sedang mengalami pandemi akibat *covid-19* . Kabupaten Musi Banyuasin merupakan salah satu daerah yang terdampak virus *covid-19* ini , per 21 April 2020 sejak pertama kalinya kasus positif *covid* di Kabupaten Musi Banyuasin hingga saat ini tanggal 23 Desember 2021 angka positif di Kabupaten Musi Banyuasin mencapai 2937 kasus positif .

Berdasarkan data yang diperoleh dari *website Covid-19 Muba* per 23 Desember 2021 terdapat total 2937 kasus positif , 2160 *suspect* dan 152 meninggal dunia. Tingginya angka kasus positif . Sehingga pemerintah Kabupaten Musi Banyuasin telah melakukan berbagai upaya untuk mengatasi wabah ini , salah satunya adalah dengan melakukan *lockdown* atau untuk sementara masyarakat dihimbau untuk menetap dirumah sebagai upaya pencegahan virus *corona*. Tetapi belum dilakukan metode khusus seperti klusterisasi yang dapat menjadi salah satu upaya untuk mencegah penyebaran virus *corona (covid-19)*.

Ada beragam cara yang dapat dilakukan untuk menangani pandemi *COVID-19* yang terus meningkat antara lain menganalisis pemberitaan, sentimen publik di social media, serta dokumen yang telah dipublikasi pemerintah terkait wabah *COVID-19*. Kecerdasan buatan (AI) dapat belajar mendeteksi 3 wabah dan memperkirakan risiko maksimum penyakit menular. *BlueDot* merupakan perusahaan yang berasal dari Kanada yang menyediakan layanan berbasis Kecerdasan buatan (AI) untuk memberi peringatan ancaman penyebaran wabah beberapa hari sebelum pusat pengendalian dan pencegahan penyakit atau organisasi kesehatan dunia mempublikasikan peringatan mereka tentang ini (Sari, 2020).

Pada penelitian ini akan dianalisis kasus *clustering K-means* 15 Kecamatan di Kabupaten Musi Banyuasin berdasarkan tingkat penyebaran kasus *COVID-19*.

*K-Means Cluster* adalah salah satu metode dan *clustering non-hierarki* yang mencoba mengelompokkan data ke dalam suatu *cluster* sehingga data dengan karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama. Ukuran kemiripan yang digunakan adalah ukuran jarak antar objek. Dua objek dengan jarak terdekat akan bergabung menjadi satu cluster. Kedekatan jarak yang dimiliki menunjukkan bahwa kedua objek tersebut memiliki tingkat kesamaan karakteristik (Sibarani, 2018).

Berdasarkan uraian tersebut maka permasalahan yang dibahas dalam penelitian ini, dapat dirumuskan bagaimana melakukan klasterisasi penyebaran *covid-19* di Kabupaten Musi Banyuasin dengan menggunakan metode *K-means Clustering*. Untuk mengetahui penyebaran kasus *COVID-19* kategori risiko sangat tinggi, tinggi, menengah, rendah dan sangat rendah di setiap Kecamatan di Kabupaten Musi Banyuasin. Beberapa aspek bisa diukur, seperti kasus positif *covid-19*, pasien *suspect* dan pasien meninggal. Diharapkan dapat mengambil keputusan yang tepat dengan tujuan menekan tingginya tingkat penyebaran *covid-19* dan dapat mengurangi angka pasien positif *covid-19* di Kabupaten MUBA.

## **1.2 Rumusan Masalah**

Berdasarkan permasalahan yang telah dibahas dapat diuraikan rumusan masalah didalam penelitian ini yaitu bagaimana penerapan metode *K-means Clustering* dalam mengelompokkan penyebaran *COVID-19* di Kabupaten Musi Banyuasin berdasarkan wilayah Kecamatan.

### 1.3 Batasan Masalah

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

1. Batasan untuk analisis *cluster K-Means* dan teori yang mendasarinya.
2. Studi kasus penelitian ini dibatasi berdasarkan data kasus *covid-19* pada 15 Kecamatan di Kabupaten Musi Banyuasin pada bulan April 2020 sampai dengan Desember 2021 berdasarkan kasus positif, *suspect*, dan pasien yang meninggal dunia..
3. Analisis *cluster* ini dilakukan dengan data yang diperoleh dari Dinas Kesehatan Musi Banyuasin dan publikasi *website* <https://covid19.mubakab.go.id/> yang mempublikasi perkembangan kasus *covid-19*.
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. Menerapkan metode *K-Means Clustering* untuk mengelompokkan kecamatan dan mengetahui karakteristik *cluster* yang terbentuk berdasarkan tingkat penyebaran kasus COVID-19 di Kabupaten Musi Banyuasin .
2. Mengetahui pengelompokkan kecamatan apakah termasuk dalam *cluster* tingkat penyebaran kasus COVID-19 sangat tinggi, tinggi, sedang, rendah atau sangat rendah.

3. Mengetahui variabel yang menjadi faktor perbedaan paling berpengaruh pada *cluster* yang terbentuk dan Kecamatan mana yang memiliki jumlah terbanyak dalam variabel tersebut.

## **1.5 Manfaat Penelitian**

Manfaat dari penelitian ini adalah sebagai salah satu upaya untuk membantu pemerintah Kabupaten Musi Banyuasin dalam mengatasi penyebaran virus *corona (covid-19)* di kabupaten Musi Banyuasin dengan melihat pengelompokan wilayah pada klasterisasi yang dilakukan pada penelitian ini. Sehingga dapat menjadi masukan bagi pemerintah dalam meningkatkan penanganan dan pengawasan kepada masyarakat khususnya di wilayah kecamatan yang memiliki risiko penularan tinggi. Mempermudah gugus tugas *covid-19* dalam mendistribusikan bantuan seperti obat – obatan kepada pasien *covid-19* di setiap *cluster*.

## **1.6 Metodologi Penelitian**

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

Tempat penelitian dilaksanakan di Dinas Kesehatan Kabupaten Musi Banyuasin yang berlokasi di Jl. Kol. Wahid Udin No.230, Serasan Jaya, Sekayu, Kabupaten Musi Banyuasin, Sumatera Selatan 30711 . Dan waktu penelitian ini dilakukan dimulai dari Agustus – 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 processor Intel® Core™ i5-9300H @ 2.40GHz*
2. *Printer*
3. *Flashdisk 32 GB*

#### b. *Software*

1. *Sistem Operasi Windows 10*
2. *Microsoft Word 2016*
3. *Microsoft Excel 2016*
4. *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*

Penelitian Klasterisasi Penyebaran Virus *Corona* di Kabupaten Musi Banyuasin dilakukan menggunakan *dataset* *Positif*, *Suspect* dan

Pasien Meninggal di Kabupaten Musi Banyuasin Tahun 2020 - 2021 dan *dataset* Rekap Harian Kasus *COVID 19* per Kecamatan di Kabupaten Musi Banyuasin. Data ini diperoleh dari situs MUBA Tanggap *Covid-19* yang dikelola oleh Pemerintah Kabupaten Musi Banyuasin. *Dataset* ini terdiri dari beberapa *file data* yang dibedakan berdasarkan waktu tercatatnya kasus per bulan, yakni sepanjang tahun 2020 -2021 .

## 2. Metode Analisa 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 yaitu dengan membersihkan data yang berupa duplikat, 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.

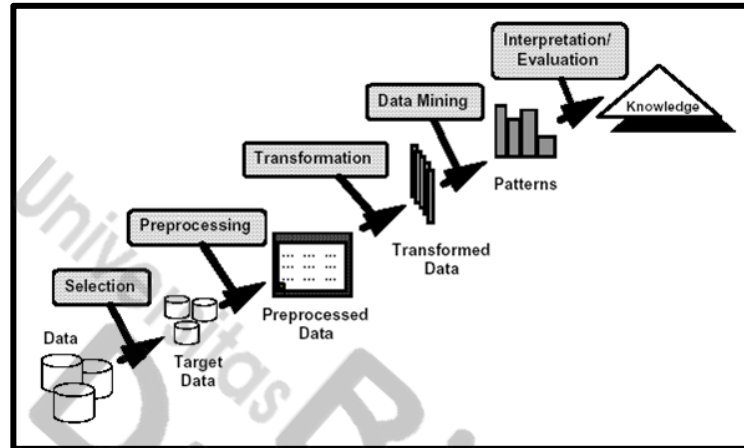
*d) Data Mining*

Tahap *Data Mining* dalam KDD 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*

Pada tahap ini pola hasil informasi dari keluaran dalam proses *data mining* perlu dirubah dalam bentuk yang mudah dimengerti oleh orang yang bersangkutan. Tahap ini II-4 mencakup pemeriksaan apakah pola atau

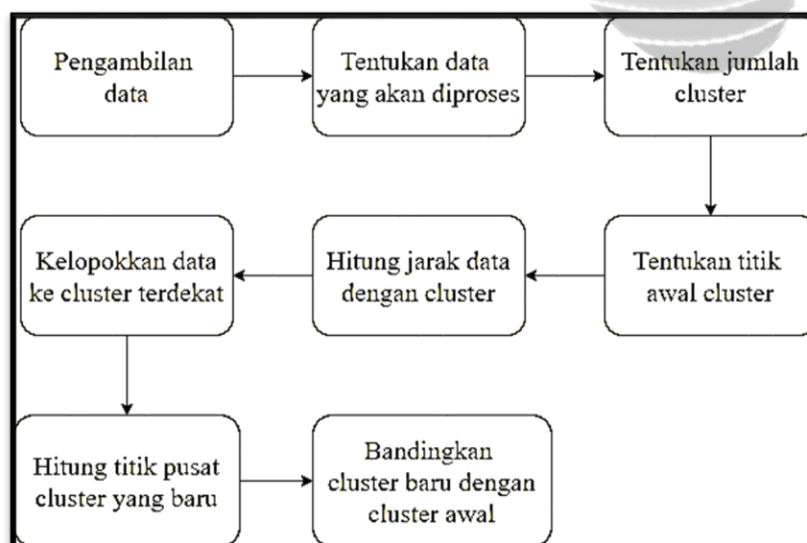
informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.



**Gambar 1. 1** Proses *Knowledge Discovery in Database* (KDD)

### 3. Penerapan Metode

Penerapan metode *K-means* pada penelitian Klusterisasi Penyebaran Virus *Corona* di Kabupaten Musi Banyuasin 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.

a) **Pengambilan Data**

Data penelitian ini diperoleh dari situs MUBA Tanggap *Covid-19*. Data yang diperoleh disesuaikan dengan spesifikasi basis data. Data tersebut kemudian disimpan ke dalam pangkalan data yang telah dibuat.

b) **Seleksi *Dataset***

Menentukan data yang ingin diproses berdasarkan Kecamatan yang ada di kabupaten Muba. *Dataset* yang digunakan bersifat kumulatif, sehingga data yang dipilih adalah total seluruh kasus sampai dengan tanggal data yang diberikan oleh Dinas kesehatan Muba.

c) **Menentukan Jumlah Klaster**

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

d) **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.

e) **Menghitung Jarak Setiap Data dengan Titik Pusat Kluster**

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

$$D_{(i,j)} = \sqrt{\sum_{k=1}^n (X_{ik} - C_{jk})^2} \quad \dots \dots \dots (1.1)$$

Yang mana dibawah ini merupakan penjelasan untuk setiap variabel pada rumus 1.1 diatas:

D = jarak kluster

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

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

n = jumlah kluster

f) **Pengelompokan Data Berdasarkan Kluster Terdekat**

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

g) **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 .

$$C_i = \frac{\sum d_i}{n_k} \dots \dots \dots (1.2)$$

h) Bandingkan Klaster Baru dengan Klaster Awal

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

**4. Metode Pengumpulan Data**

Dalam penelitian ini metode pengumpulan data yang digunakan peneliti ada 2 yaitu data primer dan data sekunder yang akan dijelaskan sebagai berikut :

**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 dinas Kesehatan maupun dengan pengurus website Muba Tanggap *covid-19* yang berhubungan dengan topik penelitian yang diangkat.

**b. Sekunder**

Literatur atau studi pustaka

Penulis mengumpulkan data dan informasi terkait dengan penelitian berupa buku cetak, 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 –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