

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. PENELITIAN SEBELUMNYA

Sebelum penelitian ini dilakukan, maka perlu dilakukan perbandingan dengan penelitian sebelumnya agar dapat menambah pemahaman yang lebih mendalam mengenai metode yang sama dan mengetahui adanya perbedaan pada objek yang diteliti. Perbandingan pada penelitian sebelumnya ini menggunakan teknik 3C2S yaitu metode yang terdiri dari membandingkan (*comparing*), mencari perbedaan (*contrasting*), memberi kritik (*criticize*), memberikan ide baru (*synthesize*), dan meringkas (*summarize*). Tabel 2.1 di bawah ini merupakan beberapa ringkasan dari penelitian sebelumnya yang terkait dengan “**Aplikasi Metode K-Means untuk Mengelompokkan Jenis Sayur Guna Mengukur Potensi Produksi Jenis Sayur Di Kabupaten Banyumas**” :

**Tabel 2.1** Penelitian Sebelumnya

No	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
1.	Penerapan <i>K-Means Cluster</i> Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara (Putrama Alkhairi, Agus Perdana Windarto)	Penelitian ini menggunakan metode yang sama yaitu metode <i>K-Means Cluster</i> untuk mengelompokkan data produksi pertanian karet.	Penelitian ini menggunakan objek pertanian karet yang diambil dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara.	Penelitian ini kurang menjelaskan dengan rinci bagaimana proses kelanjutan dari data yang telah diperoleh.	Penelitian ini dilakukan sebagai cara untuk membantu pemerintah dalam menentukan daerah pengembangan perkebunan karet dalam	Penelitian ini menghasilkan data potensi pertanian karet yang terbagi menjadi 3 <i>cluster</i> yaitu <i>cluster</i> tertinggi ada 1 kota, <i>cluster</i> menengah ada 6

No	Judul	Comparing	Contrasting	Criticize	Synthesize	Summarize
					pengembangan wilayah pertanian di provinsi Sumatera Utara.	kota, dan <i>cluster</i> terendah ada 19 kota.
2.	Penerapan Data <i>Mining</i> Dalam Mengelompokkan Produksi Daging Sapi di Pulau Sumatera Menggunakan Algoritma <i>K-Means Clustering</i> (Asril Andi Novany, M.Safii)	Penelitian ini menggunakan metode yang sama yaitu metode <i>K-Means Clustering</i> untuk mengelompokkan data produksi daging sapi.	Sumber data penelitian ini dikumpulkan berdasarkan dokumen produksi daging sapi dari Badan Pusat Statistika Nasional di pulau Sumatera, dan hanya mengelompokkan berdasarkan 2 <i>cluster</i> yaitu <i>cluster</i> tertinggi dan <i>cluster</i> terendah.	Penelitian ini hanya menggunakan 2 <i>cluster</i> untuk pengelompokan data yaitu <i>cluster</i> tertinggi dan <i>cluster</i> terendah.	Penelitian ini dilakukan sebagai cara untuk membantu pemerintahan dalam mengelompokkan jumlah produksi daging sapi berdasarkan wilayahnya.	Penelitian ini menghasilkan data produksi daging sapi yang terbagi menjadi 2 <i>cluster</i> yaitu <i>cluster</i> tertinggi ada 5 kota, dan <i>cluster</i> terendah ada 5 kota.
3.	Penerapan Data <i>Mining</i> Dalam Mengelompokkan Produksi Jagung Menurut Provinsi Menggunakan Algoritma <i>K-Means</i> (Nanda Erlangga, Solikun, Irawan)	Penelitian ini menggunakan metode yang sama yaitu metode <i>K-Means</i> untuk mengelompokkan data produksi	Penelitian ini menggunakan objek pertanian jagung yang diambil dari Badan Pusat Statistik (BPS) di tiap provinsi di	Penelitian ini hanya menggunakan 2 <i>cluster</i> untuk pengelompokan data yaitu <i>cluster</i> tertinggi dan <i>cluster</i> terendah.	Penelitian ini dilakukan sebagai cara untuk mengoptimalkan program pemerintah di bidang pertanian dan meningkatkan	Penelitian ini menghasilkan data produksi pertanian jagung yang terbagi menjadi 2 <i>cluster</i> yaitu <i>cluster</i> tertinggi ada 2 provinsi,

No	Judul	Comparing	Contrasting	Criticize	Synthesize	Summarize
		jagung.	Indonesia.		produksi jagung di Indonesia.	dan <i>cluster</i> terendah ada 32 provinsi.
4.	Penerapan Data Mining <i>K-Means Clustering</i> untuk Mengelompokkan Berbagai Jenis Merk Laptop (Putri Ulil Fatma Aulia, Sudin Saepudin)	Penelitian ini menggunakan metode yang sama yaitu metode <i>K-Means Clustering</i> untuk mengelompokkan data penjualan jenis laptop.	Penelitian ini menggunakan berbagai macam merek laptop sebagai objek penelitiannya.	Penelitian ini tidak menambahkan variabel harga pada data yang diolah, sehingga pelanggan tidak mendapatkan rekomendasi laptop yang berkualitas nemun dengan harga terjangkau.	Penelitian ini dilakukan sebagai cara untuk mempermudah konsumen dalam memilih merek laptop dengan kualitas terbaik.	Penelitian ini menghasilkan data merek laptop yang terbagi menjadi 3 <i>cluster</i> yaitu <i>cluster</i> tertinggi ada 11 item, <i>cluster</i> menengah ada 6 item, dan <i>cluster</i> terendah ada 8 item.
5.	Penerapan Data Mining Dalam Mengelompokkan Jumlah Usaha Berdasarkan Provinsi Menggunakan <i>K-Means Clustering</i> (Rahel Adelina Hutasoit, M.Safii, Iin Parlina)	Penelitian ini menggunakan metode yang sama yaitu metode <i>K-Means Clustering</i> untuk mengelompokkan data jumlah usaha.	Sumber data penelitian ini dikumpulkan berdasarkan dokumen keterangan Jumlah Usaha di Indonesia yang dihasilkan oleh Badan Pusat Statistik Nasional.	Pada penelitian ini tidak menjelaskan jenis usaha apa yang sedang diteliti.	Penelitian ini dilakukan sebagai cara untuk memberi masukan pada pemerintah agar menyediakan sarana dan prasarana yang mendukung para pengusaha di Indonesia.	Penelitian ini menghasilkan data jumlah usaha yang terbagi menjadi 3 <i>cluster</i> yaitu <i>cluster</i> tertinggi ada 4 provinsi, <i>cluster</i> menengah ada 13 provinsi, dan <i>cluster</i> terendah ada 17 provinsi.
6.	Penerapan Data Mining Dalam Mengelompokkan	Penelitian ini menggunakan	Penelitian ini menggunakan	Data pengangguran	Penelitian ini dilakukan sebagai	Penelitian ini menghasilkan data

No	Judul	Comparing	Contrasting	Criticize	Synthesize	Summarize
	Data Pengangguran Terbuka Menurut Provinsi Menggunakan Algoritma <i>K-Means</i> (Mochamad Wahyudi, Lise Pujiastuti, Solikhun)	metode yang sama yaitu metode <i>K-Means</i> untuk mengelompokan data pengangguran terbuka.	data pengangguran terbuka sebagai objek penelitiannya.	yang teliti tidak diketahui sumbernya, sehingga informasi yang disediakan kurang jelas.	cara penanganan masalah pengangguran di Indonesia oleh pemerintah.	produksi pertanian jagung yang terbagi menjadi 2 <i>cluster</i> yaitu <i>cluster</i> tertinggi ada 12 provinsi, dan <i>cluster</i> terendah ada 22 provinsi.
7.	Analisis Pesebaran Penularan Virus Corona Di Provinsi Jawa Tengah Menggunakan Metode <i>K-Means Clustering</i> (Darmansah, Ni Wayan Wardani)	Penelitian ini menggunakan metode yang sama yaitu metode <i>K-Means Clustering</i> untuk menganalisis penyebaran penularan virus corona.	Penelitian ini menggunakan objek data penularan virus corona yang diambil dari Dinas Kesehatan Provinsi Jawa Tengah.	Perhitungan metode <i>K-Means Clustering</i> pada penyebaran virus corona tidak dijelaskan dengan detail pada penelitian ini, sehingga kurang dipahami.	Hasil dari penelitian ini dapat digunakan oleh pemerintah Jawa Tengah untuk mengambil kebijakan dan keputusan pada tingkat pesebarkan penularan virus corona berdasarkan kabupaten atau kota yang ada.	Penelitian ini menghasilkan data jumlah usaha yang terbagi menjadi 3 <i>cluster</i> yaitu <i>cluster</i> tertinggi ada 16 kota, <i>cluster</i> menengah ada 1 kota, dan <i>cluster</i> terendah ada 18 kota.
8.	Data <i>Mining</i> Seleksi Siswa Berprestasi untuk Menentukan Kelas Unggulan Menggunakan Metode <i>K-Means Clustering</i>	Penelitian ini menggunakan metode yang sama yaitu metode <i>K-Means</i>	Penelitian ini menggunakan objek data siswa berprestasi yang diambil dari	Perhitungan metode <i>K-Means Clustering</i> terhadap data siswa berprestasi	Hasil dari olah data penelitian tersebut dapat memudahkan guru dalam	Penelitian ini menghasilkan data jumlah usaha yang terbagi menjadi 3 <i>cluster</i> yaitu

No	Judul	Comparing	Contrasting	Criticize	Synthesize	Summarize
	(Reyhan Putra Primanda, Aslan Alwi, Dyah Mustikasari)	<i>Clustering</i> untuk menyeleksi siswa beprestasi dan menentukan kelas unggulannya.	observasi langsung pada MTS Darul Fikri Ponorogo.	tidak dijelaskan dengan detail pada penelitian ini, sehingga kurang dipahami.	menyeleksi kategori prestasi siswa dan memangkas waktu kerja mereka.	<i>cluster</i> tertinggi ada 23 siswa, <i>cluster</i> menengah ada 25 siswa, dan <i>cluster</i> terendah ada 52 siswa.
9.	Penerapan Data <i>Mining</i> Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode <i>K-Means</i> (Genta Triyandana, Lala Aprianti Putri, Yuyun Umaidah)	Penelitian ini menggunakan metode yang sama yaitu metode <i>K-Means</i> untuk mengelompokkan data penjualan makanan dan minuman.	Penelitian ini menggunakan objek data penjualan makanan dan minuman yang diambil dari observasi langsung pada Dpom Coffee Karawang.	Proses pengolahan data yang dilakukan menggunakan tools rapidminer dalam perhitungan <i>K-Means Clustering</i> terhadap penjualan makanan dan minuman tidak dijelaskan dengan detail dan rinci, sehingga kurang dipahami.	Hasil dari penelitian ini dapat digunakan oleh pemilik Dpom Coffee sebagai bahan evaluasi dan mengatur strategi penjualan kedepannya.	Penelitian ini menghasilkan data jumlah usaha yang terbagi menjadi 3 <i>cluster</i> yaitu <i>cluster</i> tertinggi ada 7 menu, <i>cluster</i> menengah ada 40 menu, dan <i>cluster</i> terendah ada 8 menu.
10.	Penerapan Data <i>Mining</i> Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. DKI Jakarta dengan <i>K-Means</i> (Linda Maulida)	Penelitian ini menggunakan metode yang sama yaitu metode <i>K-Means</i> untuk mengelompokkan data kunjungan	Penelitian ini menggunakan objek data kunjungan wisatawan yang diambil dari Badan Pusat Statistik (BPS)	Pada penelitian ini perhitungan metode <i>K-Means Clustering</i> pada pengelompokan kunjungan wisatawan tidak dijelaskan dengan	Hasil penelitian ini dapat membantu pemerintah untuk upaya peningkatan jumlah kunjungan	Penelitian ini menghasilkan data jumlah usaha yang terbagi menjadi 3 <i>cluster</i> yaitu <i>cluster</i> tertinggi ada 1 tepat wisata, <i>cluster</i>

No	Judul	Comparing	Contrasting	Criticize	Synthesize	Summarize
		wisatawan.	di provinsi DKI Jakarta.	detail, sehingga kurang dipahami.	wisatawan pada pengenalan objek wisata dan peningkatan devisa Negara DKI Jakarta.	menengah ada 2 tempat wisata, dan <i>cluster</i> terendah ada 5 tempat wisata.
11.	<i>Data Mining Approach for Automatic Ship-route Design for Coastal Seas Using Automatic Identification System (AIS) Trajectory Clustering Analysis</i> (Daheng Zhang, Yingjun Zhang, Chuang Zhang)	Penelitian ini menggunakan konsep data <i>mining</i> dalam mengelompokkan data sama seperti penelitian yang akan dilakukan.	Penelitian ini menggunakan metode <i>Automatic Identification System (AIS) Trajectory Clustering</i> untuk melacak rute, kecepatan, dan jalur kapal.	Penjelasan penyelesaian penelitian ini kurang mendetail, sehingga masih sulit untuk dipahami.	Tujuan dari penelitian ini adalah untuk meningkatkan kecepatan dan akurasi pelatihan, dan rute otomatis yang dihasilkan memenuhi persyaratan praktik navigasi pada kapal.	Melakukan analisis terhadap jalur yang dilalui oleh kapal menggunakan data dalam jumlah besar dengan menggunakan metode <i>Automatic Identification System (AIS) Trajectory Clustering</i> .
12.	<i>Data Mining Approach to Construction Productivity Prediction for Cutter Suction Dredgers</i> (Shuo Baia, Mingchao Lia, Rui Kong, Shuai Hana, Heng Lib, Liang Qinc)	penelitian ini menggunakan konsep yang sama seperti penelitian yang akan dilakukan yaitu menggunakan konsep data <i>mining</i> dalam mengelompokkan	Metode yang digunakan untuk menganalisis produktivitas kapal keruk adalah metode <i>Random Forest, K-Nearest, Naive Bayes,</i> dan <i>eXtreme Gradient</i>	Data yang digunakan pada penelitian ini tidak diketahui sumbernya.	Tujuan dari penelitian ini adalah untuk menganalisis data pemantauan kapal keruk dan memperkirakan produktivitas yang efektif dengan melakukan pendekatan baru	Menyajikan pendekatan baru untuk menganalisis data pemantauan kapal keruk dan memperkirakan produktivitas yang efektif menggunakan metode <i>Random</i>

No	Judul	Comparing	Contrasting	Criticize	Synthesize	Summarize
		data.	<i>Enhancement</i> (XGBoost).		yang terintegrasi.	<i>Forest, K-Nearest, Naive Bayes, dan eXtreme Gradient Enhancement</i> (XGBoost)
13.	<i>Extension Data Mining Method for Improving Product Manufacturing Quality</i> (Siyuan Chena, Xingsen Lia, Renhu Liua, Shouzhen Zeng)	Penelitian ini menggunakan konsep data <i>mining</i> dalam mengelompokkan data sama seperti penelitian yang akan dilakukan.	Objek yang digunakan dalam penelitian tersebut adalah kualitas produk manufaktur.	Pada penelitian ini tidak dijelaskan jenis produk manufaktur apa yang sedang diteliti.	Penelitian ini dilakukan sebagai cara untuk kualitas produk yang dapat membantu perusahaan untuk mewujudkan manajemen kualitas yang lebih sempurna pada sistem.	Memperoleh ide baru bagi perusahaan untuk meningkatkan kualitas produk dalam mewujudkan manajemen kualitas yang lebih baik.

Tabel 2.1 menunjukkan perbandingan penelitian yang akan dilakukan dengan penelitian sebelumnya. Perbedaan penelitian sebelumnya dengan penelitian yang akan dilakukan adalah dari segi objek yang dijadikan sasaran penelitian. Pada penelitian sebelumnya belum ada yang menggunakan produksi jenis sayur di Dinas Pertanian Kabupaten Banyumas sebagai objek penelitian. Selain itu, pada penelitian ini dilakukan dengan memanfaatkan algoritma *K-Means Clustering*. Data yang diambil adalah data sayur dari hasil wawancara dengan Dinas Pertanian Kabupaten Banyumas dalam 5 tahun terakhir yaitu tahun 2018-2022 dan terdiri dari 27 Kecamatan. Data yang akan diolah dibagi menjadi 3 *cluster*, yaitu *cluster* tertinggi, *cluster* sedang, dan *cluster* terendah.

## 2.2. DASAR TEORI

Terdapat beberapa dasar teori yang dapat dijadikan sebagai pengetahuan dalam penelitian menggunakan algoritma *K-Means Clustering* untuk pengelompokan potensi produksi sayur. Dasar teori yang dapat diambil adalah sebagai berikut :

### 2.2.1. *Knowledge Discovery in Database (KDD)*

Dalam sebuah basis data KDD merupakan proses untuk menemukan pola-pola yang dapat digunakan [14]. Berikut ini adalah tahapan-tahapan yang ada dalam proses KDD, antara lain :

#### 1. *Data Selection*

Sebelum tahap penggalian informasi yang lebih mendalam, data primer dipilih untuk dilakukan pemilihan data. Data yang telah dipilih secara efektif kemudian dikumpulkan dan disimpan dalam satu dokumen untuk digunakan dalam proses *data mining*.

#### 2. *Pre-processing / Cleaning*

Proses pembersihan data utama adalah duplikat data dihapus, data yang tidak valid diperiksa, dan kesalahan pada data diperbaiki. Sebelum memulai proses *data mining*, langkah pembersihan data ini dilakukan.

### 3. *Transformation*

Proses *transformation* dalam KDD diakhiri dengan *coding* pada data. Informasi yang telah diubah akan memberikan informasi yang sesuai untuk proses *data mining*.

### 4. *Data mining*

Proses KDD berikutnya adalah *data mining* yang merupakan cara paling umum untuk mencari pola pada data dalam memilih informasi dengan menggunakan teknik tertentu. Ada berbagai macam teknik pada proses *data mining*. Pemilihan teknik yang tepat sangat bergantung pada tujuan dan proses KDD secara menyeluruh.

### 5. *Interpretation / Evaluation*

Proses KDD meliputi tahap ini, yang disebut interpretasi. Langkah ini termasuk menentukan dan memeriksa apakah pola atau informasi yang ditemukan bertentangan dengan hipotesis atau fakta sebelumnya.

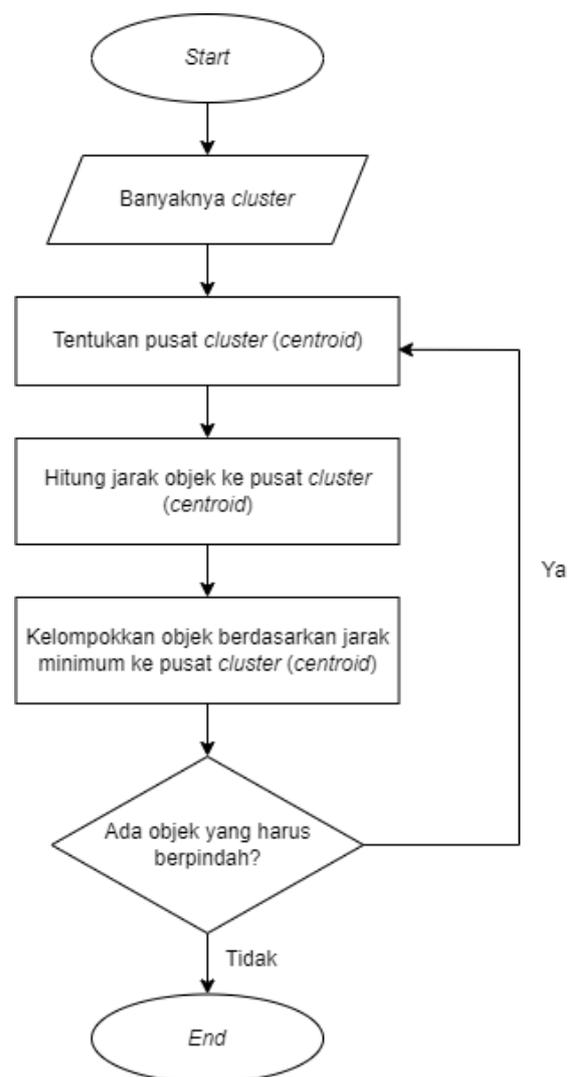
#### **2.2.2. *Data Mining***

Pengambilan sebuah keputusan berdasarkan analisis dari data yang besar biasanya menggunakan pendekatan *data mining* [13]. *Data mining* merupakan proses pengumpulan sekumpulan data dalam jumlah yang besar untuk diekstraksi sehingga menjadi informasi yang dapat digunakan untuk pengambilan keputusan [13]. Pendapat lain juga mengatakan bahwa *data mining* adalah siklus yang menggunakan prosedur statistik, numerik, penalaran buatan, dan *machine learning* untuk menghapus dan mengenali data yang bermanfaat terkait dari kumpulan data yang sangat besar [9]. Tujuan utama dari *data mining* adalah untuk menemukan, menyelidiki, atau menambang pengetahuan dari data atau informasi yang ada [9].

#### **2.2.3. *Algoritma K-Means***

Pengelompokan data adalah teknik eksplorasi dari kumpulan data yang memiliki objek dan karakteristik yang sama dikelompokkan menjadi satu untuk mengklasifikasi pengolahan data lebih lanjut [14]. *K-Means* termasuk ke dalam algoritma yang sering digunakan dalam proses pengelompokan data [9].

Pengertian algoritma *K-Means* adalah suatu algoritma yang biasanya digunakan dalam proses pengelompokan data dengan cara membagi atau memisahkan data ke kelompok yang berbeda-beda [3]. Algoritma ini mampu memperpendek jarak antara data ke *clusternya* [7]. Tujuan dari *K-Means* adalah pengelompokan data dengan memaksimalkan kemiripan data dalam satu *cluster* [13].



**Gambar 2.1** Flowchart Algoritma *K-Means*

Gambar 2.1 merupakan *flowchart* dari algoritma *K-Means*, agar proses yang dilakukan dapat tergambar dengan jelas. Berikut ini merupakan penjelasan dari *flowchart* yang telah dibuat [10] :

1. Penentuan banyaknya *cluster* yang akan dibentuk. Nilai  $K$  adalah jumlah dari *cluster*.
2. Menentukan pusat *cluster* (*Centroid*) awal. *Centroid* awal ditentukan secara acak dan jumlah *centroid* awal sama dengan jumlah *cluster*.

$$ck = \frac{\text{Jumlah dari nilai yang masuk ke dalam cluster}}{\text{Jumlah data yang masuk}} \quad (2. 1)$$

$ck$  = Pusat *cluster* (*Centroid*) awal

3. Jika *centroid* awal telah ditentukan, maka perhitungan jarak setiap data ke masing-masing *centroid* dilakukan dengan menggunakan rumus korelasi antar dua obyek yaitu *Euclidean Distance* untuk menemukan *centroid* terdekat pada masing-masing data.

$$D_{(x,y)} = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_n - Y_n)^2} \quad (2. 2)$$

$D$  = Titik Pusat

$X$  = Data *Record*

$Y$  = Data *Centroid*

4. Setelah menghitung jarak data ke *centroid*, maka langkah berikutnya adalah mengelompokkan data berdasarkan jarak minimum. Suatu data akan menjadi anggota dari suatu *cluster* yang memiliki jarak terdekat (terkecil) dari pusat *cluster*.
5. Ulangi langkah 2, lakukan perulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota *cluster* tidak berpindah ke *cluster* lain.
6. Proses akan berhenti jika tidak ada lagi perbedaan pada *cluster*.

#### 2.2.4. *Clustering*

Salah satu pengelompokkan dalam *data mining* adalah metode *clustering*. *Clustering* adalah salah satu alat bantu dalam melakukan proses *data mining* yang bertujuan untuk mengelompokkan objek ke dalam masing-masing *cluster* [11]. *Cluster* adalah metode analisis data yang sering digunakan sebagai salah satu metode dari *data mining* [12]. Tujuan *cluster* adalah untuk mengelompokkan data

dengan karakteristik yang sama ke suatu wilayah dan data dengan karakteristik yang berbeda ke wilayah yang lain [12]. Penggunaan metode *clustering* ini dapat mengklasifikasikan daerah yang padat, menemukan pola-pola distribusi secara keseluruhan, dan juga menemukan keterkaitan yang menarik antara atribut data [10].

#### **2.2.5. Sektor Pertanian**

Sektor pertanian merupakan sektor yang biasanya memproduksi bahan pangan [2]. Peranan sektor pertanian adalah sebagai salah satu sumber yang dapat menghasilkan bahan keputuhan pokok, dapat menyediakan lapangan pekerjaan bagi sebagian besar penduduk, memberikan kontribusi dalam meningkatkan pendapatan nasional, dan memberikan devisa negara yang tinggi [2]. Pengaruh besar dalam menjaga laju pertumbuhan ekonomi nasional akan terjadi jika perkembangan pada sektor pertanian terus mengalami peningkatan yang positif [3]. Sektor pertanian juga dapat dijadikan sebagai basis yang kuat dalam menciptakan pengembangan kegiatan produktivitas pedesaan melalui pengembangan usaha melalui sektor pertanian, yaitu agribisnis dan agroindustri [3].

Fungsi dari sektor pertanian meliputi [6]:

- a. Memenuhi kebutuhan pangan kepada populasi penduduk yang terus berkembang.
- b. Meningkatkan permintaan produk industri untuk mendorong pertumbuhan sektor sekunder dan tersier.
- c. Memperluas keuntungan perdagangan asing negara untuk perbaikan melalui komoditas item pertanian yang konsisten.
- d. Meningkatkan pendapatan daerah sehingga pemerintah dapat terlibat.
- e. Pengembangan lebih lanjut untuk kesejahteraan masyarakat.

#### **2.2.6. Tanaman Sayur**

Sayuran merupakan tumbuhan yang dibudidayakan oleh para petani di Indonesia untuk dijadikan sebagai bahan makanan [2]. Sayuran sangat penting

bagi manusia, karena di dalamnya mengandung berbagai sumber vitamin, mineral dan serat [2]. Bagian tumbuhan yang biasa dijadikan sebagai bahan pangan antara lain, daun contohnya bayam, batang contohnya wortel, bunga contohnya jantung pisang, buah contohnya pepaya muda, sehingga dapat dikatakan sebagian besar bagian tumbuhan dapat digunakan sebagai bahan makanan [2]. Pada umumnya masyarakat Indonesia menjadikan lahan kering sebagai tempat membudidayakan tanaman sayuran, baik tanaman utama maupun pada sistem tumpang sari [2]. Komoditas pertanian budidaya sayuran ini penting bagi ketahanan pangan nasional di Indonesia dan juga pertumbuhan ekonomi [1].

#### **2.2.7. Dinas Pertanian Kabupaten Banyumas**

Dinas Pertanian merupakan salah satu perangkat daerah di Kabupaten Banyumas, yang dibentuk berdasarkan Peraturan Daerah Nomor 7 Tahun 2019 tentang Perubahan Peraturan Daerah Kabupaten Banyumas Nomor 16 Tahun 2016 tentang Pembentukan dan Susunan Perangkat Daerah Kabupaten Banyumas, dengan tugas pokok membantu Bupati dalam perumusan kebijakan, pelaksanaan, evaluasi, dan pelaporan penyelenggaraan urusan pemerintahan di bidang pertanian dan pangan yang menjadi kewenangan daerah dan tugas pembantuan Kabupaten. Kedudukan, Rancangan Wewenang, Kewajiban, Tata Kerja, dan Tata Cara Kerja Dinas Tani Rezim Banyumas telah diatur dalam Pedoman Dinas Banyumas Nomor 81 Tahun 2020 [6].

#### **2.2.8. *RapidMiner***

*RapidMiner* adalah aplikasi yang dibuat oleh Dr. Markus Hofman dari *Institute of Technology Blanchardstown* dan *Ralf Klinkenberg* dari *rapid-i.com* dengan GUI (*Graphical UI*), sehingga dapat memudahkan penggunaanya dalam mengerjakan program kerja menggunakan aplikasi ini [12]. *RapidMiner* bersifat *open source* dan dibuat dengan menggunakan program Java dan juga dapat digunakan di sistem operasi manapun. Pada saat mengoperasikan *RapidMiner*, kemampuan *coding* tidak lagi dibutuhkan, karena *tools* yang disediakan sudah

sangat lengkap. *RapidMiner* biasanya digunakan untuk penggunaan yang sedang melakukan analisis mengenai *data mining*.

### **2.2.9. Dataset**

*Dataset* adalah bermacam-macam data yang terkait satu sama lain, diatur secara terorganisir dan biasanya diperkenalkan dalam struktur tabel [20]. Pada penelitian ini dataset yang digunakan didapatkan dari observasi langsung kepada pihak Dinas Pertanian Kabupaten Banyumas. Tujuan dari pengambilan dataset ini adalah untuk menguji sekaligus dijadikan bahan penunjang pada proses perhitungan menggunakan algoritma *K-Means Clustering* pada penelitian ini.