

BAB II

TINJAUAN PUSTAKA

2.1 *Clustering*

Clustering adalah proses mengidentifikasi kelompok objek yang memiliki kesamaan. Melalui teknik clustering, kita dapat menentukan lebih jauh kepadatan dan jarak antar wilayah objek spasial serta mengeksplorasi secara komprehensif pola sebaran dan hubungan antar atribut. Metode klasifikasi juga dapat digunakan secara efektif untuk memisahkan kelompok atau kelas objek.

Pengelompokan (*clustering*) adalah komponen dalam bidang data mining yang bersifat tanpa bimbingan (*unsupervised*). Ini adalah proses penyusunan data ke dalam kelompok atau *cluster* berdasarkan sejauh mana kesamaannya. Dalam pengelompokan, data yang memiliki kesamaan dikelompokkan bersama dalam cluster yang sama, sementara data yang tidak memiliki kesamaan ditempatkan di *cluster* yang berbeda [1].

2.2 *K-Means*

Pengertian *K-Means* yaitu algoritma pengelompokan berulang, yang membagi sekumpulan data menjadi sejumlah *K cluster* yang awalnya ditentukan. Algoritma ini mudah diimplementasikan dan dijalankan, relatif cepat, mudah diadaptasi, dan umum digunakan dalam praktik [2]. Kemudian, *K-Means* merupakan algoritma dalam teknik clustering *non-hierarki* yang berupaya mempartisi data yang ada menjadi satu atau lebih cluster [3].

Algoritma *K-Means* adalah pengelompokan data ke dalam kelompok yang memiliki karakteristik yang sama. *K-means* clustering sebagai salah satu metode data clustering non-hierarki mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* atau kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam

kelompok yang lain. Kelompok atau cluster yang didapat merupakan pengetahuan/informasi yang bermanfaat bagi pengguna kebijakan dalam proses pengambilan keputusan. Algoritma *K-means* sangat cocok digunakan untuk pengelompokan barang berdasarkan ukuran sehingga dapat lebih efektif pada saat pengiriman barang [4].

K-Means mengambil bagian dari komponen populasi untuk menjadi cluster awal. Cluster dipilih secara acak dari kumpulan populasi data. Berikutnya *K-Means* memeriksa setiap elemen pada populasi data dan menandai elemen pada salah satu pusat *cluster* yang ditentukan berdasarkan jarak minimum antar elemen ke setiap *cluster*. Posisi *cluster* akan dihitung ulang hingga seluruh elemen data terklasifikasi pada masing-masing pusat *cluster*, dan akhirnya akan terbentuk posisi pusat *cluster* yang baru. Algoritma *K-Means* pada dasarnya melakukan dua proses, yaitu proses pendeteksian lokasi pusat setiap *cluster* dan proses pencarian anggota setiap *cluster*. Berikut adalah cara kerja algoritma *K-Means*:

1. Menentukan K sebagai jumlah *cluster* yang ingin dibentuk
2. Menentukan K *centroid* (titik pusat awal) secara acak
3. Menghitung jarak setiap data yang ada ke masing-masing *centroid*
4. Kelompokkan data berdasarkan *centroid* terdekat
5. Menentukan pusat *cluster* baru
6. Kembali pada tahapan proses ke-3 menggunakan pusat *cluster* baru

Berikut adalah rumus untuk menentukan jarak data dari masing masing *centroid*:

$$[4] \quad d(P, Q) = \sqrt{\sum_{j=1}^p (P_j - Q_j)^2}$$

Keterangan:

D = Titik Dokumen

P = Data *Record*

Q = Data *Centroid*

2.3 *E-commerce*

Pengertian *E-commerce* merupakan suatu sistem transaksi jual beli yang memanfaatkan alat elektronik, seperti telepon dan internet. *E-commerce* lebih dari sekadar aktivitas jual beli produk secara daring. *E-commerce* mencakup seluruh rangkaian proses yang melibatkan pengembangan, pemasaran, penjualan, pengiriman, layanan, dan pembayaran kepada pelanggan, dengan dukungan dari jaringan mitra bisnis di seluruh dunia. Sistem *e-commerce* sangat tergantung pada sumber daya internet dan beragam teknologi informasi lainnya untuk menunjang setiap tahapan dalam proses ini [5].

E-commerce dapat membantu para pelaku bisnis dalam mempromosikan barang dagangannya secara *online*. Dengan adanya *e-commerce*, calon konsumen dapat dengan mudah membeli barang yang mereka inginkan dan memanfaatkan semua fasilitas yang tersedia. Setiap individu dapat memilih produk dengan berbagai pilihan transaksi yang mudah, beragamnya produk, atau berbelanja di toko yang menyediakan kebutuhan konsumen.

2.4 **Android**

Android adalah sistem operasi yang fungsi sebagai manajemen sumber daya perangkat keras, digunakan pada perangkat seperti ponsel, *smartphone*, dan tablet. Pada dasarnya Android merupakan platform terbuka (*Open Source*), yang memungkinkan pengembang membuat aplikasi sendiri yang dapat digunakan di banyak perangkat seluler berbeda.. Perangkat seluler pertama yang menggunakan sistem operasi Android adalah HTC *Dream* yang dirilis pada 22 Oktober 2008.. Hingga akhir tahun 2009, diperkirakan setidaknya

ada 18 jenis ponsel yang menggunakan Android. Sejak diluncurkan pada tanggal 9 Maret 2009, Android telah mengalami banyak pembaruan, termasuk versi 1.1 yang mencakup perbaikan kosmetik pada aplikasi seperti jam alarm, pencarian dengan suara, pesan Gmail, dan notifikasi *email*.

Android merupakan sistem operasi yang digunakan pada perangkat *mobile* dan berbasis Linux. Ini mencakup sistem operasi, *middleware*, dan aplikasi. Android memberikan kesempatan bagi para pengembang untuk membuat aplikasi mereka sendiri. Android telah menjadi sistem operasi yang menjalankan lebih dari satu miliar *smartphone* dan tablet [6].

2.5 Android Studio

Android Studio merupakan IDE yang secara resmi digunakan untuk mengembangkan aplikasi Android, dan didasarkan pada platform IntelliJ IDEA. Untuk mempermudah pengembangan aplikasi Android, Android Studio memanfaatkan sistem *build*, *emulator*, kode templat, dan integrasi dengan Github yang menggunakan Gradle sebagai dasarnya. Android Studio menggunakan fitur *push* untuk mengunggah perubahan kode dan sumber daya ke aplikasi yang sedang berjalan. *Editor code* membantu pengembang dengan menulis kode dan menyediakan fitur penyelesaian kode, modifikasi, dan analisis. Aplikasi yang dibuat di Android Studio kemudian dikompilasi ke dalam format APK untuk dikirim ke Google Play Store. Android Studio tersedia untuk platform desktop Mac, Windows, dan Linux.

Android Studio adalah sebuah *Integrated Development Environment (IDE)* yang dirancang untuk memfasilitasi pengembangan aplikasi Android, dan proyek ini merupakan hasil pengembangan oleh Google. Android Studio merupakan evolusi dari *Eclipse IDE* dan didasarkan pada IntelliJ IDEA, sebuah IDE Java yang sangat populer. Rencananya, Android Studio akan menjadi IDE resmi yang menggantikan *Eclipse* dalam pengembangan aplikasi Android ke depan [7].

2.6 *Java Development Kit (JDK)*

Java Development Kit adalah adalah Paket fungsi (API) untuk bahasa pemrograman Java. JDK mencakup dua komponen utama, yaitu *Java Runtime Environment (JRE)* dan *Java Virtual Machine (JVM)*. Peran utama JDK adalah untuk melakukan proses kompilasi kode Java ke dalam format bytecode yang dapat dijalankan oleh JRE (*Java Runtime Environment*). Untuk mengembangkan aplikasi berbasis Java, JDK harus diinstal di komputer pengembang, tetapi untuk menjalankan aplikasi yang sudah dibangun dengan Java, JRE adalah yang diperlukan. JRE, singkatnya, adalah komponen utama yang memungkinkan aplikasi berbasis Java berjalan pada suatu komputer setelah JRE tersebut terinstal. Tanpa JDK, pengembangan aplikasi Java tidak mungkin, sementara tanpa JRE, aplikasi Java tidak dapat dijalankan.

2.7 **Kotlin**

Kotlin adalah bahasa pemrograman yang terhubung dengan *Java Virtual Machine (JVM)*. Ini adalah bahasa pemrograman yang memiliki pendekatan praktis dalam pengembangan aplikasi Android, menggabungkan paradigma berorientasi objek dan pendekatan berbasis fungsional. Kotlin juga memiliki sifat interoperabilitas, yang memungkinkan penggunaan bahasa ini dalam proyek bersama dengan bahasa pemrograman Java. Bahasa pemrograman ini tidak hanya berlaku untuk pengembangan aplikasi Android, tetapi juga dapat diterapkan dalam pengembangan aplikasi berbasis desktop, web, dan *backend*. Kotlin pertama kali dikembangkan oleh perusahaan JetBrains, yang juga mengembangkan IntelliJ IDEA. Setelah mengalami serangkaian perkembangan, JetBrains kemudian mengumumkan bahwa Kotlin akan menjadi proyek open source, dan sejak saat itu, bahasa ini telah mengalami perkembangan yang signifikan. Google telah memberikan dukungan penuh terhadap Kotlin sebagai bahasa pemrograman untuk pengembangan aplikasi Android [8].

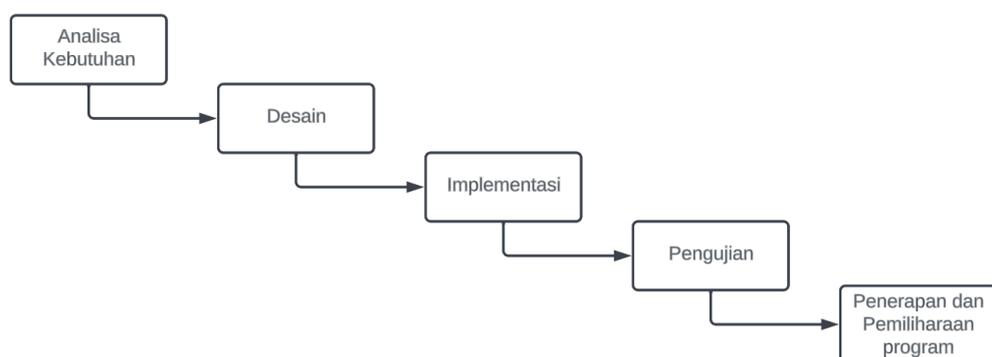
2.8 PostgreSQL

PostgreSQL adalah perangkat lunak basis data yang tersedia secara gratis berdasarkan lisensi BSD. Saat ini, perangkat lunak ini merupakan salah satu sistem basis data yang paling populer yang digunakan secara luas. PostgreSQL memberikan fitur-fitur seperti DB Mirror, PGPool, Slony, PGCluster, dan lain-lain.

PostgreSQL memiliki dukungan untuk *Standard Query Language* (SQL) yang mencakup kemampuan seperti transaksi, subquery, trigger, dan fitur lainnya. Perbedaan utama antara PostgreSQL dan sistem basis data lainnya adalah bahwa PostgreSQL memungkinkan pengguna untuk mendefinisikan fungsi SQL mereka sendiri secara independen [9].

2.9 Waterfall

Pengertian metode *waterfall* adalah suatu pendekatan yang mengikuti urutan sistematis yang dimulai dari pemahaman kebutuhan sistem, kemudian dilanjutkan ke tahap analisis, desain, pengodean, pengujian/verifikasi, dan pemeliharaan [10]. Dinamakan "*waterfall*" karena setiap tahap dalam metode ini harus menunggu selesainya tahap sebelumnya, khususnya tahap kebutuhan. Tahapan *waterfall* dapat dilihat pada gambar 2.1



Gambar 2.1 Langkah-langkah metode *waterfall*

Berdasarkan gambar 2.1 terdapat penjelasan dari setiap fase proses dari metode *waterfall* sebagai berikut:

1. Analisa Kebutuhan: Tahap ini bertujuan untuk mengidentifikasi solusi yang tepat guna untuk memenuhi kebutuhan atau mengatasi masalah dalam pengembangan perangkat lunak.
2. Desain: Pada tahap ini bertujuan membuat desain perangkat lunak yang akan dikembangkan
3. Implementasi: Tahapan ini bertujuan mengimplentasi desain dan konsep yang telah dibuat dijalankan menjadi produk atau sistem yang nyata.
4. Pengujian: Tahapan ini bertujuan untuk memastikan bahwa keseluruhan sistem bekerja dengan baik dan sesuai dengan kebutuhan pengguna
5. Pemeliharaan: Pemeliharaan dilakukan untuk memperbaiki bug yang mungkin muncul dan melakukan peningkatan atau perbaikan jika diperlukan.

2.10 Blackbox Testing

Black box testing adalah jenis pengujian yang memperlakukan perangkat lunak sebagai suatu entitas yang kinerjanya tidak diketahui secara internal. Dalam konteks ini, para penguji memandang perangkat lunak sebagai "*black box*" di mana isi atau struktur internalnya tidak dianggap penting, namun fokus pada pengujian fungsi dan interaksi eksternal. Dalam *black box testing*, perangkat lunak dieksekusi dan diuji untuk memastikan bahwa ia memenuhi kebutuhan pengguna yang telah ditetapkan pada tahap awal, tanpa perlu membongkar atau memahami detail-detail program internalnya [11]. *Black Box Testing* memusatkan perhatian pada spesifikasi fungsional perangkat lunak. Para penguji dapat menentukan serangkaian kondisi input dan melaksanakan uji coba terhadap spesifikasi fungsional program [12].

2.11 RapidMiner

Rapidminer merupakan sebuah perangkat lunak yang digunakan untuk melakukan pengolahan data mining, khususnya dalam konteks text mining. Fokus utama dari fungsi Rapidminer text mining adalah melakukan analisis

terhadap teks, di mana perangkat ini mampu mengekstrak pola-pola yang tersembunyi dalam dataset yang besar. Proses tersebut melibatkan penggabungan berbagai teknik, termasuk metode statistika, kecerdasan buatan, dan penggunaan database sebagai elemen-elemen kritis dalam menyusun informasi yang diperoleh. Dengan demikian, Rapidminer tidak hanya sekadar sebuah alat pengolahan data mining, melainkan sebuah solusi yang komprehensif untuk mengeksplorasi dan menganalisis data teks dengan menggunakan pendekatan yang terintegrasi.

2.12 *Literatur Review*

Literature review berfungsi sebagai cara untuk mengenali, mengevaluasi, dan menjelaskan semua penelitian terkait dengan pertanyaan penelitian tertentu, topik bidang, atau fenomena menarik. *Literatur review* ini dilakukan dengan cara yang realistis, melibatkan identifikasi, pemilihan, dan penilaian literatur penelitian yang relevan yang menjadi fokus diskusi [14]. Dibawah ini merupakan tabel hasil literature review sebagaimana yang dapat dilihat pada tabel 2.1 berikut:

Tabel 2.1 *Literatur Review*

NO	Literatur	Masalah	Metode	Variabel	Hasil
1	Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). Perbandingan Algoritma K- Means dan K- Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau.	Membandingkan kedua algoritma pada dataset bertujuan untuk menganalisa pola pengelompokan dan menentukan algoritma yang baik dalam pengolahan data	K-Means dan K- Medoids	Keterangan barang, jenis barang, berat barang	Pengelompokan dengan kategori transaksi tinggi, transaksi sedang dan transaksi rendah. Dalam pengolahan K- means membutuhkan waktu lebih sedikit dibandingkan K- medoids sehingga K-means lebih efesien untuk melakukan

Tabel 2.1 *Literatur Review*

NO	Literatur	Masalah	Metode	Variabel	Hasil
					pengelompokan data.
2	Dharma Putra, Y., Sudarma, M., & Swamardika, I. B. A. (2021). Clustering History Data Penjualan Menggunakan Algoritma K-Means.	Melakukan Klasifikasi data dalam proses manajemen stok dan melihat peluang dari penjualan sebelumnya yang dapat menghasilkan produk terbaik sehingga dapat menjaga kualitas produk serta kepuasan dari pelanggan	K-Means	Stok masuk, Stok keluar, stok	Pengelompokan dengan kategori sangat diminati, diminati dan kurang diminati
3	Hutabarat, S. M., & Sindar, A. (2019). Data Mining Penjualan Suku Cadang Sepeda Motor Menggunakan Algoritma K-Means.	Tidak memiliki data untuk mengetahui suku cadang yang paling banyak dibeli oleh pelanggan dalam setiap transaksi pembelian	K-Means	Stok masuk, stok keluar	Pengelompokan dengan kategori tinggi, sedang dan rendah
4	Darmi, Y. D., & Setiawan, A. (2016). Penerapan metode clustering k-means dalam	Memerlukan klasifikasi dalam memahami perilaku konsumen,	K-Means	Total penjualan, harga barang	Pengelompokan dengan kategori laku dan tidak laku

Tabel 2.1 *Literatur Review*

NO	Literatur	Masalah	Metode	Variabel	Hasil
	pengelompokan penjualan produk.	mengidentifikasi peluang produk baru dipasaran.			
5	Mudzakkir, B. D. (2018). Pengelompokan Data Penjualan Produk Pada Pt Advanta Seeds Indonesia Menggunakan Metode K-Means.	Memerlukan identifikasi dan pengkelompokan produk apa saja yang diminati sehingga bisa menyusun faktor-faktor apa saja yang dapat menarik konsumen baru untuk membeli.	K-Means	Stok barang, stok terjual	Pengelompokan dengan kategori laku dan tidak laku
6	Saritas, M. M., & Yasar, A. (2019). Performance analysis of ANN and Naive Bayes classification algorithm for data classification.	Melakukan perbandingan algoritma ANN dan Naive Bayes terhadap klasifikasi terhadap pasien kanker payudara.	ANN dan Naive Bayes	Biomarkers. Glucose, Resistin, Age, BMI, HOMA, Leptin, Insulin, Adiponectin, MCP-1	Dalam penelitian ini, mengklasifikasikan pasien kanker payudara dengan kategori sehat dan penderita. Dan didapatkan hasil akurasi 86,95 dengan ANN dan 83,54 dengan algoritma Naive Bayes.
7	Dody Firmansyah, M., Saddam, S., Sfenrianto, S., Bachtiar, S., & Kaburuan, E. R. (2019). Sales Information System Using Web	Melakukan pengembangan bisnis dengan cara membuat sistem penjualan berbasis web.	-	-	Menghasilkan sistem penjualan berbasis web yang dapat dijalankan oleh para konsumen.

Tabel 2.1 *Literatur Review*

NO	Literatur	Masalah	Metode	Variabel	Hasil
	for Small Business (Case Study: CV. Tanaka Service).				
8	Soegoto, D. S., & Cica, C. (2018, August). Design of Web-based Sales Information System on Fashion Shop in Bandung, Indonesia.	Membuat konsep bisnis modern dengan menggunakan sistem penjualan fashion, agar lebih mudah dan dapat diakses dimanapun.	-	-	Menghasilkan sistem penjualan berbasis web yang dapat dijalankan oleh para konsumen untuk melakukan transaksi pembelian.
9	Hutabarat, J. C. J. (2016). Perancangan aplikasi menentukan berat badan ideal dengan menggunakan algoritma K-Means clustering.	Membuat aplikasi menentukan berat badan ideal dengan menerapkan algoritma K-Means	K-Means	Tinggi badan, Lingkar lengan, Berat Badan	Menghasilkan aplikasi berbasis desktop dengan mengelompokan data menjadi kategori Kurus, Normal, serta Gemuk

Penelitian yang dilakukan Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). Dalam mengklasifikasikan data transaksi bongkar muat di Provinsi Riau menghasilkan kesimpulan bahwa Dalam pengolahan K-means membutuhkan waktu lebih sedikit dibandingkan K-medoids sehingga K-means lebih efisien untuk melakukan pengelompokan data.

Pada penelitian Dharma Putra, Y., Sudarma, M., & Swamardika, I. B. A. (2021). Dalam mengklasifikasi data penjualan menggunakan algoritma K-Means untuk mendapatkan data tentang barang yang sangat diminati, diminati dan kurang diminati.

Pada penelitian Hutabarat, S. M., & Sindar, A. (2019). Dalam mengklasifikasikan data Penjualan Suku Cadang Sepeda Motor menggunakan algoritma K-Means untuk mendapatkan data Suku Cadang dengan tingkat penjualan tinggi, sedang dan rendah.

Pada penelitian Darmi, Y. D., & Setiawan, A. (2016). Dalam memahami perilaku konsumen dengan cara melakukan klasifikasi dengan menggunakan algoritma K-Means untuk mendapatkan data produk yang laku dan tidak laku.

Pada penelitian Mudzakkir, B. D. (2018). Dalam mengklasifikasikan data Penjualan Pt Advanta Seeds Indonesia untuk mendaptkan data produk yang laku dan tidak laku.

Pada penelitian Saritas, M. M., & Yasar, A. (2019). Melakukan perbandingan algoritma ANN dan Naive Bayes terhadap klasifikasi terhadap pasien kanker payudara dan menghasilkan hasil akurasi 86,95 dengan ANN dan 83,54 dengan algoritma Naïve Bayes.

Selanjutnya penelitian yang dilakukan oleh Dody Firmansyah, M., Saddam, S., Sfenrianto, S., Bachtiar, S., & Kaburuan (2019) dalam penelitian tersebut menghasilkan sistem penjualan barang elektronik di CV. Tanaka Service berbasis website. Akan tetapi sistem yang dibangun tidak menggunakan algoritma untuk melakukan klasifikasi data dan penggunaan sistem penjualan berbasis *website* sudah tidak relevan dengan keadaan sekarang dimana para konsumen lebih sering melakukan transaksi pembelian dengan aplikasi mobile dibandingkan dengan menggunakan website.

Pada penelitian yang dilakukan Soegoto, D. S., & Cica, C. (2018) dalam penelitian tersebut menghasilkan sistem penjualan fashion berbasis website. Akan tetapi pada aplikasinya pembeli tidak dapat melakukan pembayaran dan sistem yang dibangun tidak menggunakan algoritma untuk melakukan klasifikasi data dan penggunaan sistem penjualan berbasis *website* sudah tidak relevan dengan keadaan sekarang dimana para konsumen lebih sering melakukan transaksi pembelian dengan aplikasi mobile dibandingkan dengan menggunakan website.

Dan yang terakhir pada penelitian yang dilakukan oleh Hutabarat, J. C. J. (2016). Dalam membuat aplikasi dengan menerapkan algoritma K-Means menghasilkan aplikasi berbasis desktop yang dapat menghitung dan mengelompokan data berat badan ideal menjadi kategori Kurus, Normal, serta Gemuk.

Dengan demikian, algoritma K-Means ini cocok untuk digunakan dalam melakukan klasifikasi, sehingga penulis memutuskan untuk mengadopsi arsitektur ini dalam penelitiannya setelah merujuk beberapa penelitian sebelumnya. Penelitian yang akan dilakukan oleh penulis berjudul "Pembangunan Aplikasi Mobile Menggunakan Klasifikasi K-Means untuk Penjualan Peralatan dan Mesin Pertanian".