

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 PT. Tunas Dwipa Matra**

Pada PT. Tunas Dwipa Matra berada di Jl. Raya Batanghari No.444, Selo Rejo, Batanghari, Kabupaten Lampung Timur. PT. Tunas Dwipa Matra, sebuah perusahaan besar yang bergerak di bidang distribusi utama sepeda motor Honda. Perusahaan ini berdiri pada tanggal 23 Januari 1978 di Indonesia.

Dalam perkembangannya, PT. Tunas Dwipa Matra bergerak di jaringan jasa penjualan, perawatan, perbaikan, dan suku cadang sepeda motor Honda di Indonesia dengan total jumlah jaringan 76 outlet penjualan dan outlet perawatan yang tersebar di seluruh Indonesia.

#### **2.2 Data Set**

Data set merupakan Kumpulan data yang diatur dalam format yang terstruktur, seperti tabel atau file, dan berisi informasi data berbagai sumber. Data set dapat berupa numerik, teks, gambar atau gabungan dari semuanya.

#### **2.3 Data Mining**

Data *mining* adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan (Suntoro, 2019). Jiawei Han, yang merupakan salah satu ahli terkemuka dalam bidang data *mining*, telah mengelompokkan jenis data *mining* menjadi beberapa kategori berdasarkan fokus analisis. Berikut adalah beberapa kategori data *mining* menurut Jiawei Han:

- *Clustering* (Pengelompokan)

Ini adalah kategori yang fokus pada pengelompokan data ke dalam kelompok-kelompok yang memiliki karakteristik serupa. Tujuannya adalah untuk menemukan struktur dalam data yang mungkin tidak

diketahui sebelumnya. Algoritma *K-means* dan *Hierarchical Clustering* adalah contoh dari teknik dalam kategori ini.

- *Classification* (Klasifikasi)

Klasifikasi adalah kategori yang fokus pada memberikan label atau kategori pada data berdasarkan ciri-ciri tertentu. Tujuannya adalah untuk memprediksi atau mengklasifikasikan data ke dalam kelompok tertentu. Algoritma seperti Naive Bayes, *Decision Trees*, dan *Support Vector Machines* termasuk dalam kategori ini.

- *Association Rule Mining* (Pencarian Aturan Asosiasi)

Kategori ini berfokus pada menemukan hubungan atau asosiasi antara item atau Atribut dalam data. Tujuannya adalah untuk mengidentifikasi pola atau aturan yang sering muncul dalam data. Algoritma Apriori adalah salah satu contoh dalam kategori ini.

- *Anomaly Detection* (Deteksi Anomali)

Ini adalah kategori yang fokus pada mendeteksi data yang tidak sesuai dengan pola mayoritas dalam dataset. Tujuannya adalah untuk menemukan data yang merupakan pengecualian atau anomali dalam data. Algoritma dalam kategori ini dapat digunakan untuk mendeteksi *fraud*, kesalahan data, atau insiden tak terduga lainnya.

- *Regression* (Regresi)

Regresi adalah kategori yang fokus pada memprediksi nilai numerik berdasarkan data yang telah ada. Ini digunakan untuk analisis data berbasis numerik, seperti peramalan penjualan atau estimasi harga saham.

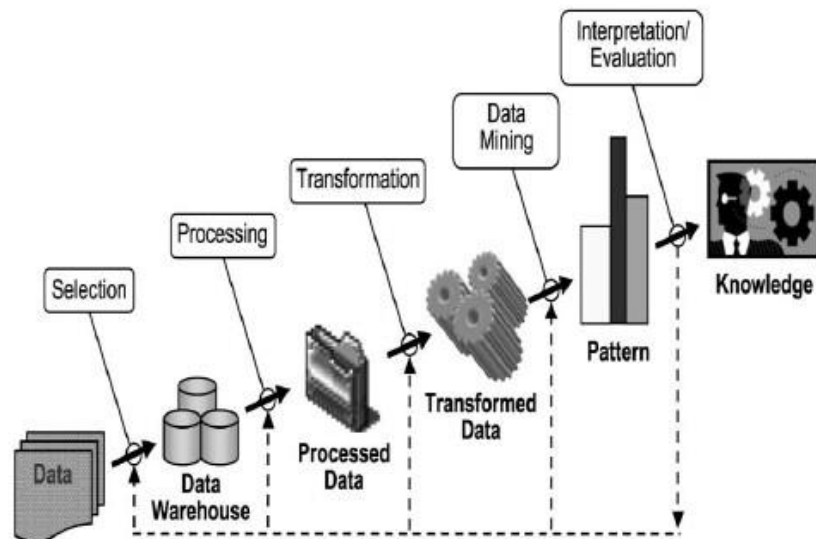
- *Summarization* (Peringkasan)

Kategori ini berfokus pada membuat ringkasan data yang besar dan kompleks. Tujuannya adalah untuk menghasilkan ringkasan data

yang dapat membantu pemahaman dan pengambilan keputusan lebih cepat. Contoh dalam kategori ini termasuk teknik seperti peringkasan teks atau peringkasan dokumen.

### 2.3.1 Tahapan *Data Mining*

Agar data yang ada dapat digunakan dan digali, data tersebut harus dipersiapkan terlebih dahulu agar dalam proses penggalian informasi tidak terjadi kesalahan. Terdapat beberapa teknik dalam mempersiapkan sebuah data yang akan digunakan dalam proses penggalian (Amna et al., 2023), diantara tahapan tersebut adalah



Gambar 2.1 Tahapan *Data Mining*

- a) *Data selection* Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses *Data Mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.
- b) *Pre-processing / Cleaning* Sebelum proses *Data Mining* dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses *cleaning* mencakup

antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

- c) *Transformation* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *Data Mining*. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.
- d) *Data Mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *Data Mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
- e) *Interpretation / Evaluation* Pola informasi yang dihasilkan dari proses data *mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

## **2.4 Clustering**

Pengelompokan atau *clustering* adalah melakukan pemisahan / pemecahan / segmentasi data ke dalam sejumlah kelompok (*cluster*) menurut karakteristik tertentu yang diinginkan. Dalam pekerjaan pengelompokan, label dari data belum diketahui dan dengan pengelompokan diharapkan dapat diketahui kelompok data untuk kemudian diberi label sesuai keinginan. *Clustering* merupakan *unsupervised learning* yang membagi data menjadi beberapa kelompok berdasarkan kemiripan Atribut yang dimiliki masing-masing data

tersebut. Awalnya karakteristik tiap kelompok tidak didefinisikan terlebih dahulu, namun karakteristik tersebut kemudian akan terlihat berdasarkan homogenitas karakter data dalam suatu kelompok yang akan berbeda dengan kelompok lainnya(Hardiani, 2022).

Analisis *Cluster* merupakan suatu teknik analisa *multivariate* untuk mencari dan mengorganisir informasi tentang variabel sehingga secara relatif dapat dikelompokkan dalam kelompok yang homogen atau “*cluster*” dapat dibentuk. *Cluster* dibentuk dengan metode kedekatan yang secara internal harus homogen (anggota adalah serupa untuk satu sama lain) dan sangat secara eksternal tak sejenis (anggotanya tidak seperti anggota dari *cluster* yang lain). Analisis *cluster* dapat menerima suatu data masukan yang beragam. Ini biasanya disebut pengukuran “kesamaan”, dapat juga disebut “kedekatannya”, dan “kemiripannya”. Beberapa ahli merekomendasikan penggunaan standardisasi data, *cluster* dapat dihitung dalam skala yang berbeda dan standardisasi akan memberi pengukuran dengan menggunakan unit yang berbeda. Tujuan pekerjaan pengelompokan (*clustering*) data dapat dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan pengelompokan untuk penggunaan. Jika tujuannya untuk pemahaman, kelompok yang terbentuk harus menangkap struktur alami data, biasanya proses pengelompokan dalam tujuan ini 5 hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan inti seperti peringkasan atau *summarization* (rata-rata, standar deviasi), pelabelan kelas pada setiap kelompok untuk kemudian digunakan sebagai data latih klasifikasi dan sebagainya. Sementara jika untuk penggunaan, tujuan utama pengelompokan biasanya adalah mencari *prototype* kelompok yang paling representative terhadap data, memberikan abstraksi dari setiap objek data dalam kelompok di mana sebuah data terletak di dalamnya(Alodia et al., 2021).

## 2.5 Algoritma *K-means Clustering*

Salah satu metode Data Mining yang sering digunakan yaitu Algoritma *K-means Clustering*, untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada pengelompokan data yang memiliki kemiripan sehingga dengan pengelompokan tersebut dapat dilakukan analisis. Data -data dipilih menjadi beberapa kelompok dengan kriteria yang telah ditentukan lalu dikumpulkan menjadi satu dalam sebuah *cluster*, di mana setiap *cluster* ini memiliki *Centroid* atau titik pusat (Suntoro, 2019). Berikut adalah tahapan-tahapan untuk menggunakan Algoritma *K-means*:

### 2.5.1 Penentuan *Cluster* Awal

Dalam menentukan  $n$  buah pusat *cluster* awal dilakukan pembangkitan bilangan random yang merepresentasikan urutan data input. Pusat awal *cluster* didapatkan dari data sendiri bukan dengan menentukan titik baru, yaitu dengan random pusat awal dari data.

### 2.5.2 Perhitungan Jarak Dengan Pusat *Cluster*

Untuk mengukur jarak antar data dengan pusat dengan *cluster* digunakan *euclidian distance*, algoritma perhitungan jarak data dengan pusat *cluster*:

- a. Pilih nilai data dan nilai pusat *cluster*
- b. Hitung *euclidian distance* data dengan tiap pusat *cluster*

$$d(x_i, \mu_j) = \sqrt{(x_i - \mu_j)^2} \dots \dots \dots (1)$$

Penjelasan :

$x_i$  : Data kriteria

$\mu_j$  : *Centroid* pada *cluster* ke  $j$

### 2.5.3 Pengelompokan Data

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini

menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat *cluster* terdekat. Adapun cara pengelompokan data tersebut adalah :

- a) Pilih nilai jarak tiap pusat *cluster* dengan data.
- b) Cari nilai jarak terkecil.
- c) Kelompokkan data dengan pusat *cluster* yang memiliki jarak terkecil.

#### 2.5.4 Penentuan Pusat *Cluster* Baru

Untuk mendapatkan pusat *cluster* baru bisa dihitung dari rata-rata nilai anggota *cluster* dan pusat *cluster*. Pusat *cluster* yang baru digunakan untuk melakukan Iterasi selanjutnya, jika hasil yang didapatkan belum konvergen. Proses Iterasi akan berhenti jika telah memenuhi maksimum Iterasi yang dimasukkan oleh user atau hasil yang dicapai sudah konvergen (pusat *cluster* baru sama dengan pusat *cluster* lama). Algoritma penentuan pusat *cluster* :

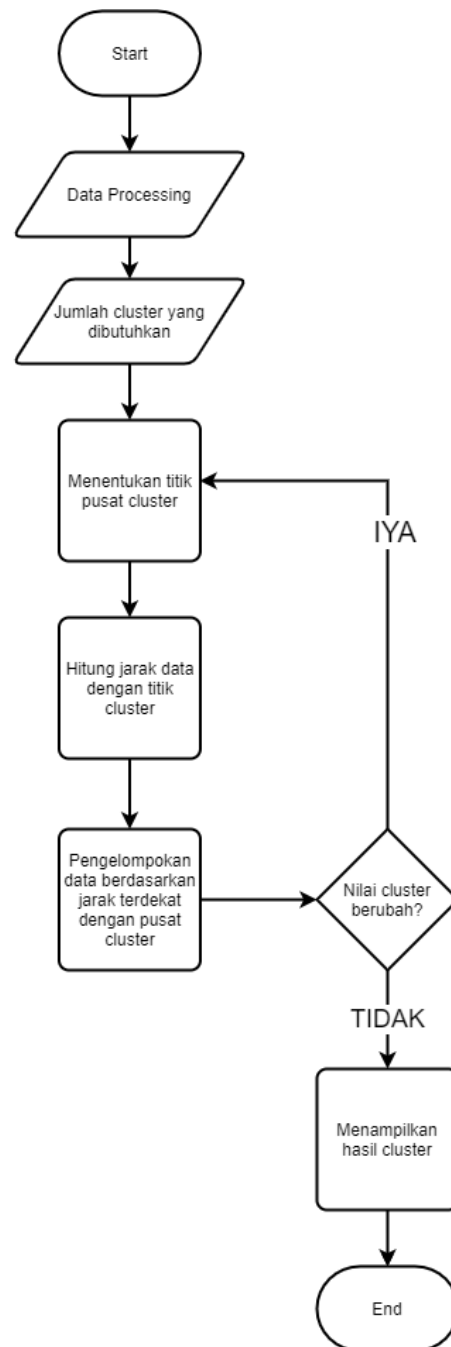
- a) Cari jumlah anggota tiap *cluster*
- b) Hitung pusat baru dengan rumus

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in S_j} x_j \dots \dots \dots (2)$$

Penjelasan :

$\mu_j(t+1)$  : *Centroid* baru pada Iterasi ke 1

$N_{sj}$  : Banyak data pada *cluster* sj



Gambar 2.2 *Flowchart* Algoritma *K-means*

## 2.6 Segmentasi

Segmentasi adalah proses pembagian pasar ke dalam kelompok pembeli yang memiliki kebutuhan, karakteristik, atau perilaku serupa yang dapat memengaruhi keputusan mereka untuk membeli. Setiap kelompok ini, atau segmen, membutuhkan strategi pemasaran yang berbeda dan perhatian yang khusus (Prakasawati et al., 2019).



### 2.6.1 Jenis Segmentasi

Ada beberapa jenis segmentasi

1. Segmentasi Demografis

Segmentasi ini membagi pasar berdasarkan karakteristik demografis seperti, usia, jenis kelamin, pendapatan, dan pendidikan.

2. Segmentasi Geografis

Segmentasi ini membagi pasar berdasarkan lokasi geografis seperti negara wilayah, dan kota.

3. Segmentasi Psikografis

Segmentasi ini membagi pasar berdasarkan karakteristik psikologis seperti nilai-nilai, kepercayaan, dan gaya hidup.

4. Segmentasi Perilaku

Segmentasi ini membagi pasar berdasarkan perilaku pembelian seperti frekuensi pembelian, jumlah pembeli, dan merek yang digunakan

5. Segmentasi Psikografis—Behavioral

Segmentasi ini menggabungkan faktor psikologis dan perilaku untuk membagi pasar menjadi kelompok-kelompok yang lebih terperinci.

### 2.6.2 Tujuan Segmentasi

Adapun beberapa tujuan dari segmentasi yaitu:

1. Meningkatkan Efektivitas Pemasaran

Segmentasi ini membantu perusahaan untuk memahami karakteristik dan kebutuhan masing-masing kelompok pelanggan, mengembangkan produk dan layanan yang relevan dan menarik bagi setiap kelompok.

## 2. Memperluas Pangsa Pasar

Perusahaan dapat mengidentifikasi kelompok pelanggan yang belum terlayani dengan baik oleh pesaing atau yang memiliki kebutuhan khusus yang belum terpenuhi

## 3. Meningkatkan Keuntungan

Dengan menargetkan kelompok pelanggan yang paling menguntungkan, perusahaan dapat meningkatkan keuntungan dan mengoptimalkan alokasi sumber daya pemasaran.

## 4. Meningkatkan Loyalitas Pelanggan

Dengan memahami kebutuhan dan preferensi pelanggan, perusahaan dapat mengembangkan produk dan layanan yang lebih relevan dan menarik bagi setiap kelompok.

## 5. Mengurangi Resiko

Segmentasi pasar dapat perlu membantu perusahaan mengurangi resiko dan meningkatkan keberhasilan kampanye pemasaran. Dengan memahami karakteristik dan kebutuhan masing – masing kelompok pelanggan, perusahaan dapat menghindari mengeluarkan biaya untuk kampanye pemasaran yang tidak efektif dan meminimalkan resiko kegagalan kampanye pemasaran

### **2.7 Rapid Miner**

*Rapid Miner* adalah sebuah *software* yang digunakan untuk pengolahan *Data Mining*. *Rapid Miner* mengolah dengan cara mengekstrak pola dari *Dataset* dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database. Tujuan dari penggunaan *Rapid Miner* adalah untuk mendapatkan informasi bermutu tertinggi dari teks yang diolah. *Rapid Miner* digunakan pada berbagai penelitian untuk pengujian data yang di teliti (Mesran et al., 2020).

## 2.8 Penelitian Terdahulu

Berikut adalah beberapa ringkasan dari beberapa penelitian sebelumnya yang terkait dengan penelitian ini:

Tabel 2.1 Penelitian Terdahulu

No	Judul	Metode	Tujuan	Peneliti	Tahun penelitian	Hasil
1	“penerapan Algoritma K-Medoids untuk menentukan segmentasi pelanggan	Algoritma K-Medoids <i>Clustering</i>	Tujuan dari penelitian ini adalah untuk mengetahui tipe dan karakteristik pelanggan untuk memberikan rekomendasi strategi pemasaran untuk perusahaan	Anggi Ayu Dwi Sukistyawati, Mujiono Sadikin	7 september, 2021	Hasil dari penelitian ini adalah data yang siap untuk dilakukan pengolahan data, mengetahui berapa banyak pelanggan yang melakukan transaksi ditahun 2019, dan menormalisasikan berdasarkan data yang sudah di <i>clustering</i>

2	<p><i>K-means Clustering</i> untuk segmentasi produk berdasarkan analisis Recency, Recency, Recency (RFM) pada Data Transaksi Penjualan</p>	<p>Algoritma <i>K-means Clustering</i> dan RFM</p>	<p>Tujuan utama dari peneliti ini adalah untuk segmentasi produk dalam mengkombinasikan model analisis FRM dan <i>Clustering</i> menggunakan metode <i>K-means Clustering</i>.</p>	<p>Maulidah Fithriyah, Muhammad Ainul Yakin, SyahiduzZaman</p>	<p>31 agustus, 2021</p>	<p>Peneliti mendapatkan hasil 3 klompok produk yang berbeda, dengan nilai recency, recency yang cukup baik. Dan mengetahui klompok produk yang banyak diminati serta memiliki persediaan untuk produk-produk</p>
3	<p>Segmentasi konsumen berdasarkan Model Recency, Frequency, Monetary dengan metode <i>K-means</i></p>	<p>RFM, <i>K-means Clustering</i></p>	<p>Penelitian ini bertujuan untuk mengidentifikasi pelanggan dalam fase pertama yaitu CRM dan perusahaan dapat</p>	<p>Atik Febriani, Syahfarah Ashari Putri</p>	<p>24 juli, 2020</p>	<p>Penelitian ini menghasilkan 3 klastee yang terbentuk dengan menggunakan metode <i>K-means Clustering</i></p>

			memahami hal yang diinginkan pelanggan.			g dengan menggunakan SPSS.
4	3 Analisis Segmentasi Pelanggan Ritel Z Menggunakan <i>K-means Clustering</i>	Algoritma K_Means, <i>Cluster</i> , Ritel	Peneliti bertujuan untuk mengidentifikasi karakteristik pelanggan lainnya dan menghubungkan dengan objek yang sebelumnya tidak diketahui perusahaan.	Roaida Yanti, Prita Nurkhailisa Maradjabessy	30 agustus, 2023	Hasil analisis segmentasi dari penelitian ini ialah telah didapatkan dan diketahui <i>cluster</i> yang paling potensial dilihat dari karakteristik yang dimiliki.

## 2.9 Visualisasi Data

Visualisasi Data adalah proses penyajian data dalam bentuk grafik yang membuat informasi mudah dimengerti, hal ini membantu menjelaskan tentang fakta dan menentukan arah tindakan. Definisi visualisasi data menjelaskan tentang pentingnya data dengan menempatkan data dalam konteks visual. Hal ini melibatkan penciptaan dan studi representasi visual dari data yang dikenal sebagai informasi. Visualisasi data memungkinkan pengguna untuk mempermudah dalam membaca Kumpulan dari sebuah data. Visualisasi dapat dilakukan dengan menggunakan dashboard, di mana teks, pola, dan korelasi yang tidak terdeteksi dapat dengan mudah divisualisasikan dengan menggunakan perangkat lunak visualisasi (Fithri et al., 2022).

Visualisasi data tidak hanya mengubah data menjadi grafik visual, akan tetapi visualisasi data juga memerlukan perencanaan. Setiap jenis data memerlukan teknik visualisasi yang sesuai berdasarkan kebutuhannya. Berdasarkan tingkat kompleksitas data, untuk menghasilkan solusi yang berharga perlu melibatkan berbagai disiplin ilmu, seperti statistika, *Data Mining*, desain grafis, dan *Information Visualization*.