

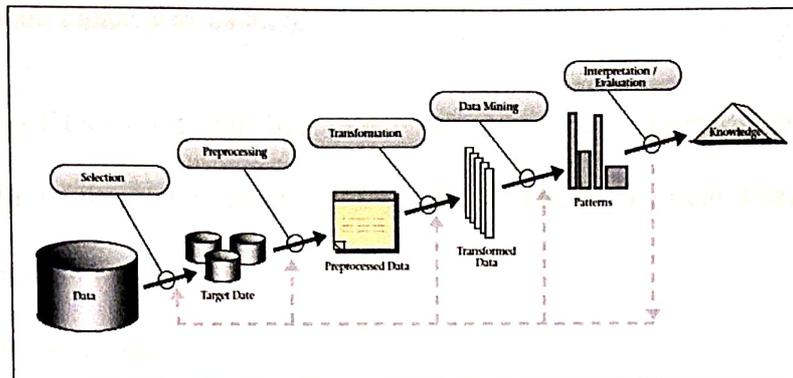
## BAB II

### LANDASAN TEORI

#### 2.1 Knowledge Discovery In Database

*Knowledge Discovery In Database (KDD)*, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menentukan keteraturan, pola atau hubungan dalam sebuah set data yang berukuran besar. Keluaran dari data mining banyak digunakan untuk pengambilan keputusan dimasa depan [1]. *Knowledge Discovery in Database (KDD)* merupakan sebuah terknik yang digunakan untuk membentuk pola atau rule dalam informasi. Informasi yang dihasilkan bersumber dari sebuah data yang berukuran besar yang biasa dikenal dengan tambang data yang biasanya disimpan dalam database yang dimana awalnya belum diketahui dan kemudian menghasilkan data yang bermanfaat [2]

Data dengan nilai yang hilang dan data berlebih dihapus, dikompilasi ke dalam sebuah tabel yang digunakan sebagai data pelanggan akhir. Data yang terintegrasi memerlukan pemilihan atribut untuk memilih data yang relevan sesuai dengan kebutuhan yang ingin dicapai. Langkah selanjutnya dalam pembersihan data adalah transformasi data. Pada langkah ini, Anda harus mengubah data yang telah dibersihkan terlebih dahulu. Data yang diubah dapat diproses menggunakan data *mining*.



**Gambar 2.1** Knowledge Discovery In Database (KDD)

(Sumber : KDD Proses)[3].

Menurut Fayyad et al. (1996), KDD didefinisikan sebagai "proses non-trivial untuk mengidentifikasi pola yang valid, baru, bermanfaat, dan dapat dipahami dari data yang besar." Proses ini melibatkan berbagai tahapan yang saling terkait untuk memperoleh informasi yang bermakna dari data mentah.

Dalam konteks ini, "non-trivial" berarti bahwa proses tersebut tidak sederhana dan membutuhkan teknik analitis yang canggih. Pola yang dihasilkan harus "valid," yang berarti benar secara statistik dan dapat diandalkan. Pola tersebut juga harus "baru," mencerminkan pengetahuan yang sebelumnya tidak diketahui. Akhirnya, pengetahuan yang diperoleh harus "bermanfaat" untuk pengambilan keputusan dan "dapat dipahami" oleh pengguna akhir.

### 2.1.1 Tahapan dalam Proses KDD

Proses KDD terdiri dari beberapa tahap utama, yang semuanya memiliki peran penting dalam menghasilkan pengetahuan yang bermanfaat. Berikut adalah tahapan-tahapan tersebut:

#### 1. Data Selection

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.

#### 2. Preprocessing

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning dengan tujuan untuk membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Juga dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal

#### 3. Transformation

Yaitu proses coding 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 database

#### 4. Data Mining

Proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu

## 5. 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 dengan *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau *hipotesa* yang ada sebelumnya atau tidak. Data *Mining* digunakan untuk *ekstraksi informasi* penting yang tersembunyi dari dataset yang besar. Dengan adanya Data Mining maka akan didapatkan suatu permata berupa pengetahuan di dalam kumpulan data – data yang banyak jumlahnya[4].

## 2.2 Data Mining

Data *mining* adalah proses yang menggunakan pembelajaran mesin untuk *menganalisis* dan *mengekstraksi* data secara otomatis.(sumber1) Data Mining adalah proses penggalian informasi dan pola yang bermanfaat dari suatu information yang sangat besar. Proses *information mining* terdiri dari pengumpulan *information*, *ekstraksi information*, *analisa information*, dan *statistik information*. Ia juga umum dikenal sebagai *information discovery*, *information extraction*, *information/sample analysis*, *statistics harvesting*, dan lainnya. Empat proses dalam *information mining* ini akan menghasilkan model/ pengetahuan yang sangat berguna. *information mining* dapat didefinisikan sebagai penguraian kompleks dari sekumpulan *information* menjadi informasi yang memiliki potensi secara implisit (tidak nyata/jelas) yang sebelumnya belum

diketahui. Ia juga dapat didefinisikan sebagai penggalian dan analisis dengan menggunakan peralatan otomatis atau semi otomatis, dari sebagian besar information yang memiliki tujuan yaitu menemukan pola yang memiliki arti atau maksud. Data mining termasuk ke dalam *information discovery* di dalam *database* (KDD)[5].

Data *Mining* merupakan sebuah inti dari proses KDD, meliputi dugaan algoritma yang mengeksplor data, membangun model dan menemukan pola yang belum diketahui bersifat otomatis, dapat didefinisikan sebagai pengorganisasian proses untuk pengidentifikasian yang benar, berguna dan penemuan pola dari kumpulan data yang besar dan kompleks. [6]

### 2.3 Naive Bayes

*Naive Bayes* merupakan *algoritma* yang umum digunakan dalam penambangan data karena memberikan hasil klasifikasi yang efektif meskipun tidak ada hubungan langsung antara atribut. *Algoritma* tersebut masih menunjukkan kinerja yang baik dan kompetitif dalam proses *klasifikasi*. *Metode Naive Bayes* dapat diterapkan secara efektif dalam sistem rekomendasi. Kesederhanaan dan efisiensinya memungkinkan *skalabilitas* yang baik bahkan dengan kumpulan data yang besar, sehingga ideal untuk rekomendasi [7].

*Algoritma Naive Bayes* dapat melakukan *proses klasifikasi* data menggunakan pemodelan *statistik*, dan merupakan salah satu metode *algoritma*

*teknologi klasifikasi* yang tercepat dan termudah untuk digunakan. Metode ini memungkinkan Anda menghitung nilai *probabilitas* untuk hasil data uji dari peristiwa data. *Naive Bayes* adalah teknik pembelajaran mesin yang menggunakan perhitungan *probabilitas*. Ide dasar yang digunakan adalah *Teorema Bayes*, yaitu menghitung nilai probabilitas untuk melakukan klasifikasi. *Teorema Bayes* adalah metode probabilitas dan statistik untuk memperkirakan probabilitas masa depan berdasarkan pengalaman masa lalu [8].

*Naive Bayes* adalah metode *klasifikasi*. NBC dipilih karena merupakan metode *klasifikasi* yang sederhana dan efisien. *Naive Bayes* dapat diterapkan pada kumpulan data yang cukup besar dan dapat menangani data yang tidak lengkap (termasuk nilai yang hilang). *Teori Naive Bayes* memiliki kemampuan klasifikasi yang sebanding dengan pohon keputusan dan jaringan saraf. Bahkan *algoritma Naive Bayes* mencapai akurasi dan kecepatan tinggi ketika diterapkan pada basis data yang berisi sejumlah besar data. *Teorema Bayes* dapat diungkapkan dalam Persamaan [9].

*Teorema Bayes* adalah rumus dasar yang digunakan dalam *algoritma Naive Bayes* untuk menghitung *probabilitas* suatu kategori C berdasarkan fitur X. Rumusnya adalah sebagai berikut:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Rumus Menghitung Probabilitas

Dimana:

1.  $P(C|X)$  adalah probabilitas kategori C diberikan fitur X,
2.  $P(X|C)$  adalah probabilitas fitur X dalam kategori C,
3.  $P(C)$  adalah probabilitas awal kategori C,
4.  $P(X)$  adalah probabilitas fitur X.

Dalam penerapan *Naive Bayes*, asumsi independensi fitur digunakan, yang memungkinkan penyederhanaan rumus menjadi:

$$P(X|C) = P(x_1|C) \cdot P(x_2|C) \cdots P(x_n|C)$$

Rumus Penyederhanaan

Dengan menggunakan rumus ini, *Naive Bayes* dapat menghitung probabilitas untuk kategori produk terlaris berdasarkan data yang ada dan memprediksi kategori produk yang paling relevan.

## 2.4 Klasifikasi

Klasifikasi (*Classification*), merupakan langkah dalam menemukan satu model yang dapat membedakan kelas data, tujuannya ialah bisa memperkirakan suatu kelas dari labelnya yang tidak dikenal sebelumnya. Yang termasuk dalam komponen klasifikasi diantara *Neural Network*, *Decision Tree*, *k-Nearest Neighbor*, dan *Naive Bayes* [10].

Klasifikasi adalah proses menempatkan data yang baru ke dalam kategori atau kelas yang sudah ditetapkan sebelumnya. Proses klasifikasi melibatkan

penggunaan variabel kategori untuk menentukan target klasifikasi. Misalnya, dalam konteks klasifikasi keuntungan, data dapat dikelompokkan ke dalam tiga kategori, yaitu kecil, menengah, dan besar. Klasifikasi adalah teknik untuk mengkategorikan atau mengelompokkan data berdasarkan kelas atau label yang dimilikinya. *Algoritma* yang digunakan dalam klasifikasi termasuk dalam kategori supervised learning. Teknik ini juga dapat dipakai untuk memprediksi kecenderungan atau pola data di masa depan[7].

Klasifikasi merupakan salah satu teknik dalam pengolahan data yang bekerja dengan cara objek yang dipergunakan dibagi menjadi kelas-kelas dengan jumlah kelas sesuai dengan yang diinginkan. Klasifikasi dapat menciptakan suatu pola yang dapat memisahkan tiap-tiap kelas data yang bertujuan guna menentukan objek yang tergolong ke dalam kategori tertentu dilihat dari perilaku serta atribut dari kelompok yang telah didefinisikan. Klasifikasi yang dilakukan dalam penelitian ini bertujuan untuk menggolongkan data termasuk ke dalam kelas blogger professional dan blogger amatir. Salah satu penelitian sebelumnya tentang Klasifikasi Penelitian Klasifikasi data stunting di Kabupaten/Kota di Indonesia berdasarkan faktor penyebab stunting pada balita, yaitu menggunakan metode clustering dengan *algoritma K-Means*. Tujuannya adalah untuk membantu pemerintah dalam mengambil kebijakan yang sesuai terkait penurunan prevalensi stunting pada balita berdasarkan karakteristik dan permasalahan masing-masing *cluster*. Hasil penelitian menunjukkan bahwa dengan bantuan metode *elbow* menghasilkan 2 *cluster* sebagai cluster terbaik dengan nilai *selisih Sum of Square Error(SSE)* sebesar 1401.5156, dimana cluster 1 merupakan cluster dengan faktor

penyebab stunting tinggi yang terdiri dari 324 kabupaten/kota, dan cluster 2 merupakan cluster dengan faktor penyebab stunting rendah yang terdiri dari 49 kabupaten/kota. Data latih dan data uji yang digunakan untuk klasifikasi berjumlah 75 data latih dan 25 data uji. Berdasarkan master pelanggan yang dijadikan data latih, telah berhasil mengklasifikasikan 23 data dari 25 data yang diuji. Sehingga berhasil memprediksi pelanggannya dengan nilai precision mencapai 100%, nilai recall mencapai 91%, nilai accuracy mencapai 92%[9].

#### **2.4.1 Naive Bayes dalam Klasifikasi Produk Terlaris di PetShop**

*Naive Bayes* merupakan salah satu algoritma yang efektif digunakan untuk klasifikasi produk terlaris di PetShop karena algoritma ini mampu memberikan hasil yang cepat dan akurat dengan memanfaatkan probabilitas fitur-fitur yang relevan terhadap suatu kelas tertentu. Pada kasus ini, kelas yang dimaksud adalah kategori produk terlaris. *Aplikasi Naive Bayes* dapat membantu manajemen PetShop dalam mengidentifikasi produk yang paling diminati pelanggan, sehingga dapat mengoptimalkan pengelolaan stok dan strategi pemasaran.

#### **2.4.2 Proses Klasifikasi dengan Naive Bayes**

##### **1. Preprocessing Data**

Data historis penjualan diproses terlebih dahulu untuk menghilangkan nilai-nilai kosong dan outlier yang dapat memengaruhi hasil klasifikasi. Selain itu, data kategori (seperti jenis produk dan merek) dikonversi ke dalam bentuk numerik atau one-hot encoding agar dapat digunakan oleh algoritma.

## 2. Perhitungan Probabilitas Prior

Probabilitas prior untuk setiap kelas, yaitu produk terlaris dan tidak terlaris, dihitung berdasarkan proporsi jumlah produk yang termasuk dalam masing-masing kelas terhadap total data.

## 3. Perhitungan Probabilitas Kondisional

Probabilitas *kondisional* dihitung untuk setiap fitur pada kelas tertentu. Misalnya, probabilitas bahwa suatu produk dengan jenis makanan termasuk dalam kelas produk terlaris dihitung menggunakan frekuensi kemunculannya pada data historis.

## 4. Prediksi Kelas

Naive Bayes menghitung probabilitas posterior menggunakan formula:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(C)}$$

Rumus Menghitung Probabilitas posterior

Produk akan diklasifikasikan sebagai terlaris jika probabilitas posterior untuk kelas "laris" lebih tinggi dibandingkan kelas "tidak terlaris".

## 5. Akurasi

dalam algoritma *Naive Bayes* merupakan metrik evaluasi yang mengukur seberapa sering model memprediksi label kelas dengan benar. Rumus akurasi secara umum adalah:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

## 6. Precision

adalah salah satu metrik evaluasi dalam klasifikasi, terutama penting ketika fokus utama adalah mengukur seberapa akurat model dalam memprediksi kelas positif. Precision menjawab pertanyaan: Dari semua prediksi positif yang dibuat oleh model, berapa banyak yang benar-benar positif?

$$Precision = \frac{TP}{TP+FP}$$

## 7. Recall

adalah salah satu metrik evaluasi dalam klasifikasi yang mengukur seberapa baik model mengidentifikasi semua data positif. *Recall* menjawab pertanyaan: Dari semua data yang seharusnya diklasifikasikan sebagai positif, berapa banyak yang berhasil ditemukan oleh model?

$$Recall = \frac{TP}{TP+FN}$$

## 8. F1-Score

adalah metrik evaluasi dalam klasifikasi yang menggabungkan *Precision* dan *Recall* ke dalam satu angka harmonis. *Metrik* ini berguna ketika ada ketidakseimbangan kelas, sehingga kita perlu mempertimbangkan baik *False Positive* (FP) maupun *False Negative* (FN) secara bersamaan:

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

## 1. $P(C | X)$ – Probabilitas Posterior

Probabilitas bahwa data  $X$  termasuk dalam kelas  $C$  setelah mempertimbangkan bukti (fitur-fitur dari data  $X$ ).

Dalam konteks klasifikasi produk terlaris di PetShop,  $P(C|X)$  adalah probabilitas bahwa suatu produk tertentu akan menjadi produk terlaris berdasarkan fitur-fitur seperti jenis produk, harga, merek, dan sebagainya.

$$P(C_i) = \frac{\text{Jumlah sampel pada kelas } C_i}{\text{Jumlah total sampel}}$$

### Rumus Probabilitas

- $P(C_i)$  adalah probabilitas prior dari kelas  $C_i$ , yaitu probabilitas bahwa sebuah sampel berasal dari kelas  $C_i$  sebelum mempertimbangkan atribut atau informasi tambahan.
- Pembilang: Jumlah sampel pada kelas  $C_i$  / Jumlah sampel pada kelas  $C$  / Jumlah sampel pada kelas  $C_i$ . Menyatakan banyaknya sampel yang termasuk dalam kelas  $C_i$ .
- Penyebut: Jumlah total sampel / Jumlah total sampel. Menyatakan jumlah seluruh sampel yang ada dalam dataset.

### 3. $P(C)$ Prior Probability

Probabilitas awal dari kelas  $C$  sebelum mempertimbangkan data  $X$ . Prior merepresentasikan frekuensi relatif kelas tersebut dalam dataset. Misalnya,  $P(C)$  adalah probabilitas bahwa suatu produk adalah produk terlaris tanpa melihat fitur-fitur spesifiknya terlebih dahulu.

$$P(C_i) = \frac{\text{Jumlah sampel pada kelas } C_i}{\text{Jumlah total sampel}}$$

#### *Rumus Prior Probability*

- $P(C_i)$  = Probabilitas prior untuk kelas  $C_i$
- Jumlah sampel pada kelas  $C_i$  = Banyaknya sampel dalam kelas tertentu  $C_i$
- Jumlah total sampel = Total seluruh sampel dalam dataset

### 4. $P(X)$ Evidence (Probabilitas Total)

Probabilitas total dari data  $X$  terjadi, terlepas dari kelasnya. Evidence merupakan faktor normalisasi agar probabilitas posterior  $P(C|X)$  berada dalam rentang  $[0, 1]$ . Probabilitas ini dapat dihitung sebagai:

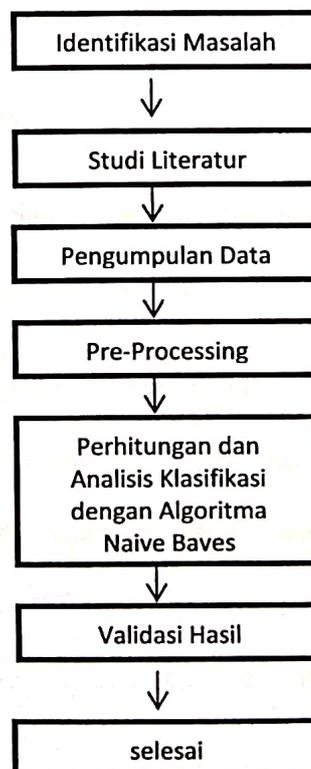
$$P(X) = \sum_{i=1}^n P(X|C_i) \cdot P(C_i)$$

#### Rumus Probabilitas Total

## 2.5 Metode Penelitian

Kerangka kerja (framework) diperlukan untuk membantu dalam penyusunan penelitian ini agar jelas tahapannya. Kerangka kerja ini merupakan langkah-langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas [11].

Penelitian ini dilakukan pada Bulan Januari di Nanda Chaniago Petshop, yang beralamat di Jalan Lintas Sumatera, Kota Raja, Kampung Pajak. Tahapan penelitian melibatkan pemilihan data yang diperoleh langsung dari Owner Nanda Chaniago Petshop untuk diperoleh lebih lanjut dalam pemilihan atribut yang akan diperoleh *preprocessing*.



**Gambar 2.2** Kerangka kerja penelitian

Berdasarkan kerangka kerja penelitian yang telah digambarkan , maka dapat diuraikan pembahasan masing-masing tahap dalam penelitian adalah sebagai berikut :

### 1. Identifikasi Masalah

Pada tahap ini penulis melakukan identifikasi masalah terhadap klasifikasi penjualan pada Nanda Chaniago Petshop. Identifikasi masalah ini bertujuan untuk mengkaji permasalahan yang ada pada Nanda Chaniago Petshop.

### 2. Studi Literatur

Pada tahap ini penulis mengumpulkan sejumlah buku-buku, dan jurnal untuk menambah wawasan yang berkaitan dengan masalah dan tujuan penelitian, yang dimana dapat membantu mempelajari dan memahami teori pada penelitian.

### 3. Pengumpulan Data

Pada tahap pengumpulan Data, penulis melakukan pengamatan langsung kelapangan dan melakukan wawancara pada owner Nanda Chaniago Petshop.

#### a. Pengamatan Langsung (Observation)

Penelitian dengan metode perception ini dilakukan dengan melakukan pengamatan langsung terhadap objek yang akan diteliti yang bertujuan untuk memperkuat information/data, mengetahui serta mendapatkan informasi secara langsung mengenai penjualan Pada toko Nanda Chaniago Petshop.

#### b. Wawancara (Interview)

Wawancara merupakan tahap dimana penulis melakukan pengajuan pertanyaan dan meminta jawaban secara langsung antara narasumber dan pewawancara. Tujuan dari wawancara adalah untuk mendapatkan informasi dimana sang pewawancara mengajukan pertanyaan-pertanyaan untuk dijawab secara langsung oleh narasumber/owner Nanda Chaniago Petshop.

#### 4. Pre-Processing

Pada tahap ini peneliti melakukan persiapan *information* penjualan makanan kucing pada Nanda Chaniago Petshop. Setelah itu peneliti melakukan seleksi atribut yang berpengaruh dalam *classification*.

##### a. Data Seleksi

Pemilihan (seleksi) *information* baru sekumpulan *information* operasional perlu dilakukan sebelum tahap penggalian informasi dalam data mining dimulai. *Information* hasil seleksi yang akan digunakan untuk proses data mining. Disimpan dalam suatu berkas, terpisah dari operasional.

##### b. Transformation

Pada tahap ini penulis melakukan *transformasi information* yaitu proses perubahan informasi ke dalam format yang sesuai untuk diproses dalam data mining. *Information* di rubah dengan *organize comma seperated files* (CSV) atau *Andrew's Ridiculos arrange* (ARFF) sebelum bisa diaplikasikan ke Rapidminer. Data tersebut berupa Data penjualan makana kucing yang telah di format pada

tahap sebelumnya yang ditransformasikan yaitu mengubah format yang awalnya berupa format dari *excel*.

## 5. Perhitungan dan Analisis dengan Algoritma Naïve Bayes

Pada tahap ini penulis melakukan perhitungan *Naïve Bayes* dengan menggunakan *Aplikasi Rapidminer* terhadap informasi data yang telah didapatkan dari Nanda Chaniago Petshop.

## 6. Validasi Hasil

Pada tahap ini penulis melakukan validasi hasil untuk menunjukkan kedekatan hasil pengukuran dengan nilai *Accuracy, Precision dan Review*.

## 7. Pembuatan Laporan

Setelah semua tahapan penelitian dilakukan, maka penulis akan membuat laporan sebagai dokumentasi penelitian agar dapat dimanfaatkan pada waktu yang akan datang.

## 2.6 Objek Penelitian

### 2.6.1 Petshop

Petshop merupakan kelinik hewan yang bergerak di bidang jasa layanan seperti pengobatan, grooming, penitipan hewan dan juga penjualan berbagai jenis makanan, obat – obatan. Aksesoris hewan serta breeding hewan, Ketika memelihara hewan ada beberapa hal yang harus diperhatikan, sebab hewan juga memerlukan perawatan yang baik agar dapat tumbuh dengan baik dan sehat. Hewan peliharaan harus diperhatikan kondisi kesehatannya, makanannya,

kandangannya, dan juga kebersihan. Kebersihan menjadi salah satu hal terpenting yang harus di perhatikan karena kebersihan hewan peliharaan kita dapat menghindarinya dari berbagai penyakit yang di sebabkan oleh virus dan kuman, Kondisi kebersihan hewan jika tidak diperhatikan maka hewan peliharaan akan mati. Beberapa penyakit yang di derita hewan peliharaan dapat juga menimbulkan efek yang berbahaya bagi manusia. Kondisi kebersihan hewan juga mempengaruhi kondisi hewan peliharaan kita, jika tidak diperhatikan maka hewan peliharaan akan mati. Perawatan untuk menjaga kebersihan dan kesehatan hewan peliharaan dapat dilakukan dengan cara memandikan, memangkas bulu dan kuku, pemberian obat, dan lain-lain. Perawatan hewan peliharaan sering di sebut Grooming. Groom dalam kamus Bahasa Inggris Indonesia artinya adalah mengurus, merawat, rapi atau pelihara. Pemilik hewan peliharaan dapat menggunakan teknik Grooming untuk membantu mempertahankan dan meningkatkan kesehatan hewan peliharaan yang di milikinya

Memprediksi produk apa yang akan menjadi terlaris di petshop memiliki dampak besar terhadap keberhasilan bisnis. Dengan mengetahui produk yang paling diminati oleh pelanggan, pemilik petshop dapat mengatur stok dengan lebih baik, meningkatkan penjualan, mengurangi risiko over stock, serta meningkatkan kepuasan pelanggan[12].



**Gambar 2.3 Toko Nanda Chaniago Petshop**

## **2.7 Rapidmier**

Rapidminer adalah paket alat pembelajaran mesin praktis yang dirancang untuk penelitian pendidikan dan berbagai aplikasi. Rapidminer dapat memecahkan masalah penambangan data dunia nyata, terutama klasifikasi, yang menjadi dasar untuk pendekatan pembelajaran mesin. Perangkat lunak ini ditulis dalam hierarki Java menggunakan metode berorientasi objek dan dapat berjalan di hampir semua platform sistem operasi. Rapidminer sangat mudah digunakan dan dapat digunakan pada banyak level. Rapidminer mengimplementasikan algoritma pembelajaran yang dapat diterapkan pada data tabular dari baris perintah.

Rapidminer menyertakan alat untuk pra-proses data, klasifikasi, pengelompokan, regresi, asosiasi, dan visualisasi. Rapidminer memproses data, memasukkannya ke dalam skema pembelajaran, dan dapat menganalisis pengklasifikasi yang dihasilkan beserta kinerjanya. Dan semua ini dapat dilakukan

tanpa menulis kode program apa pun. Contoh penggunaan Rapidminer meliputi penerapan metode pembelajaran pada kumpulan data dan menganalisis hasilnya untuk mendapatkan informasi tentang data, atau penerapan beberapa metode dan membandingkan kinerjanya untuk membuat pilihan[13].



**Gambar 2.4** Aplikasi Rapidminer

## 2.8 Penelitian Terdahulu

### 2.8.1 Tabel Penelitian Terdahulu

Referensi Penelitian 1	
Judul	Implementasi Data Mining Untuk Memprediksi Penjualan Produk Terlaris Pada Petshop Menggunakan Algoritma Naive Bayes
Nama Penulis	Andi Maslan
Tahun	2024
Hasil	Penelitian ini memprediksi penjualan produk di Berry Catszone Batam Petshop menggunakan algoritma Naive Bayes. Penghitungan dan pengolahan data dilakukan secara manual menggunakan Microsoft Excel serta dengan bantuan perangkat lunak RapidMiner, menghasilkan nilai akurasi sebagai hasil penelitian. Dalam penelitian ini, algoritma Naive Bayes telah berhasil diimplementasikan untuk memprediksi penjualan produk terlaris di Berry Catszone Batam Petshop. Penggunaan algoritma ini memungkinkan untuk melakukan klasifikasi dan analisis data penjualan dengan tingkat akurasi yang baik. Berdasarkan hasil pengujian, nilai akurasi dari prediksi yang dilakukan mencapai 90.41%. Selain itu, class precision untuk klasifikasi prediksi "laris" adalah 88.24%,

	<p>sementara class precision untuk klasifikasi prediksi "tidak laris" adalah 92.30%. Class recall untuk klasifikasi prediksi "laris" adalah 90.90%, dan class recall untuk klasifikasi prediksi "tidak laris" adalah 92.31%.</p>
--	--

### 2.8.2 Tabel Penelitian Terdahulu

Referensi Penelitian 2	
Judul	Implementasi Data Mining Untuk Mengklasifikasikan Produk Terlaris Pada Koperasi AL-BAEDLOWI Menggunakan Algoritma Naive Bayes
Nama Penulis	Irfaniyyah
Tahun	2025
Hasil	Berdasarkan penelitian dalam klasifikasi data penjualan produk di Koperasi Al-Baedlowi menggunakan algoritma Naive Bayes dengan metode penelitian Knowledge Discovery in Database (KDD) didapatkan bahwa hasil pengujian model mendapat accuracy sebesar 80.00%, precision kelas positif Kurang Laris sebesar 83.33%, dan recall kelas positif Kurang Laris sebesar 83.33%, di mana dari sebanyak 10 data pengujian terbagi menjadi sejumlah 4 untuk data berlabel Laris dengan prediksi

	<p>dari model yaitu 3 benar dan 1 salah, dan sejumlah 6 untuk data berlabel Kurang Laris dengan prediksi yaitu 5 benar dan 1 salah. Selain itu didapatkan nilai AUC sebesar 0.917 untuk kelas positif Kurang Laris yang berarti model yang telah dibuat berdasarkan kriteria nilai AUC dapat dikategorikan sebagai klasifikasi yang sangat baik (Excelent Classification).</p>
--	--

### 2.8.3 Tabel Penelitian Terdahulu

Referensi Penelitian 3	
Judul	Penerapan Data Mining Untuk Klasifikasi Produk Terlaris Menggunakan Algoritma Naive Bayes Pada Bengkel Motor
Nama Penulis	Alvin Julianto, Sri Andayani
Tahun	2022
Hasil	Penentuan produk laris dan tidak laris dengan menggunakan metode klasifikasi dan algoritma naïve bayes pada aplikasi rapid miner diperoleh bahwa algoritma naïve bayes dapat dipakai untuk melakukan penelitian terkait produk lairs dan tidak laris pada bengkel motor Aldo Motor.