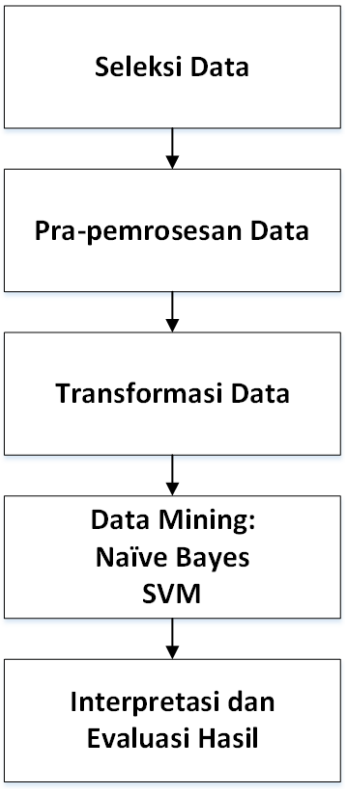
BAB III

ANALISA DAN PERANCANGAN

3.1 Arsitektur Sistem

Arsitektur sistem yang dirancang dalam penelitian ini bertujuan untuk mendukung analisis data menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) dalam menentukan minat pembeli smartphone. Arsitektur ini mencakup proses pengumpulan data, pra-pemrosesan data, penerapan algoritma, evaluasi model, hingga perbandingan hasil. Diagram arsitektur sistem dapat dilihat pada Gambar 3.1.



Gambar 3.1 Arsitektur Sistem

Tahapan-tahapan pada Gambar 3.1 dijelaskan sebagai berikut:

1. Seleksi Data

   Pada tahap ini, data yang relevan dipilih berdasarkan kebutuhan analisis.

   a. Sumber Data adalah Data penjualan Erafone Suzuya Baru Rantauprapat yang berisi informasi jumlah penjualan smartphone berdasarkan merek selama 2 tahun (2023-2024).

   b. Variabel yang tersedia:

      1. Bulan (periode transaksi)
      2. Apple, Asus, Infinix, Oppo, Realme, Samsung, Tecno, Vivo, Xiaomi (jumlah unit terjual per merek)
      3. Jumlah (total unit terjual dalam satu bulan)
      4. Omset (Rp) (total pendapatan penjualan setiap bulan)

2. Pra-Pemrosesan Data

   Pra-pemrosesan data bertujuan untuk membersihkan dan mempersiapkan data sebelum digunakan dalam analisis. Adapun langkah-langkah Preprocessing:

   a. Pembersihan Data (Data *Cleaning*)

      membersihkan data dari *missing values* dan menghapus data duplikat.

   b. Pemeriksaan *Outliers*

      memeriksa apakah ada nilai ekstrem pada jumlah penjualan dan omset.

   c. Pemeriksaan Konsistensi Data

      memastikan bahwa jumlah penjualan per bulan sesuai dengan total unit terjual di kolom "Jumlah".

3. Transformasi Data

   Pada tahap ini, data dikonversi ke format yang lebih sesuai untuk model prediksi. Adapun langkah-langkah transformasi data adalah:

   1. Normalisasi Data (*Scaling* Data)
      a. Variabel yang memiliki rentang nilai yang besar dibanding variabel lainnya.
      b. Dilakukan *Min-Max* Normalization agar skala lebih seragam dengan rumus:

$$Xmin = \frac{X - X\ min}{Xmax - Xmin}$$

   2. Diskretisasi Omset (Kategori Omset)

      Omset dikategorikan menjadi tiga kelas yaitu Omset Tinggi, Omset Sedang, dan Omset Rendah.

   3. Penambahan Variabel Baru yang Bernama Minat Pembeli yang terdiri dari Tinggi, Sedang, dan Rendah.

4. Penerapan Algoritma *Naïve Bayes* dan SVM

   Tahap ini bertujuan untuk membangun model prediksi berdasarkan data yang telah diproses. Algoritma *Naïve Bayes* digunakan untuk menghitung probabilitas setiap kelas berdasarkan distribusi data. Prediksi dilakukan berdasarkan probabilitas tertinggi. Algoritma *Support Vector Machine* (SVM) digunakan untuk membuat *hyperplane* optimal untuk memisahkan kategori omset dengan menggunakan *kernel trick* jika data tidak bisa dipisahkan secara linear. Proses *training* dan *testing model* menerapkan pembagian data menjadi Training Data (70%), dan Testing Data (30%). Kemudian model diuji menggunakan akurasi, *precision*, *recall*, dan *F1-score*.

5. Interpretasi dan Evaluasi Hasil

   Pada tahap ini, hasil dari model dianalisis untuk mendapatkan wawasan bisnis. Mengukur seberapa tepat model memprediksi kategori omset berdasarkan jumlah penjualan. Melakukan perbandingan hasil *Naïve Bayes* vs SVM berdasarkan matrik akurasi, *precision*, *recall*, dan *F1-score*.

3.1.1 Seleksi Data

Seleksi Data adalah tahap awal dalam proses *Knowledge Discovery in Database* (KDD) yang bertujuan untuk memilih atribut yang paling relevan dalam mempengaruhi penjualan dan omset, menentukan variabel mana yang akan digunakan dalam model prediksi, dan menghilangkan variabel yang tidak berkontribusi signifikan terhadap analisis.

A. Sumber Data yang Digunakan

   Dataset ini berisi informasi penjualan smartphone di Erafone Suzuya Baru Rantauprapat sepanjang tahun 2023-2024.

   1. Jumlah Data

      24 bulan (Januari 2023 – Desember 2024).

   2. Jumlah Variabel

      7 atribut (Penjualan 5 Merek, Jumlah, Omset).

Tabel 3.1 Penjualan Smartphone Erafone Suzuya Rantauprapat Tahun 2023-2024

Bulan	Merk Smartphone					Jumlah	Omset
	Apple	Oppo	Samsung	Vivo	Xiaomi		
Jan-23	30	10	25	1	7	73	Rp475.564.869
Feb-23	12	2	5	1	2	22	Rp164.754.955
Mar-23	24	16	23	3	4	70	Rp423.029.238
Apr-23	21	27	31	2	13	94	Rp493.834.237
May-23	30	14	40	1	10	95	Rp640.004.503
Jun-23	37	20	41	6	7	111	Rp632.426.390
Jul-23	34	16	35	3	3	91	Rp611.809.900
Aug-23	61	21	36	3	10	131	Rp858.485.574
Sep-23	42	23	32	4	7	108	Rp594.121.785
Oct-23	42	18	34	4	6	104	Rp604.620.894
Nov-23	35	18	53	2	6	114	Rp657.228.384
Dec-23	70	23	59	7	9	168	Rp1.165.377.930
Jan-24	66	9	53	4	5	137	Rp934.906.000
Feb-24	46	7	46	6	12	117	Rp843.497.000
Mar-24	30	16	49	1	9	105	Rp581.283.000
Apr-24	51	22	54	0	11	138	Rp842.663.000
May-24	39	26	39	3	13	120	Rp740.782.000
Jun-24	37	17	44	0	11	109	Rp650.306.000
Jul-24	25	31	58	0	11	125	Rp630.513.000
Aug-24	25	29	56	1	2	113	Rp704.838.000
Sep-24	17	23	52	3	5	100	Rp654.228.000
Oct-24	14	20	51	4	9	98	Rp543.245.000
Nov-24	32	29	41	2	6	110	Rp639.764.000
Dec-24	44	39	52	1	8	144	Rp816.124.000

Sumber: **Data penjualan Erafone Suzuya Mall Rantauprapat**

B. Penentuan Atribut

Berikut adalah atribut yang terdapat dalam dataset beserta keterangannya:

Tabel 3.2 Atribut Dataset

No.	Nama Atribut	Tipe Data	Deskripsi	Dipilih
1	Bulan	Kategorikal	Periode waktu transaksi (Januari 2023-Desember 2024)	Tidak
2	Apple	Numerik	Jumlah unit smartphone Apple yang terjual dalam bulan tersebut	Ya
3	Oppo	Numerik	Jumlah unit smartphone Oppo yang terjual dalam bulan tersebut	Ya
4	Samsung	Numerik	Jumlah unit smartphone Samsung yang terjual dalam bulan tersebut	Ya
5	Vivo	Numerik	Jumlah unit smartphone Vivo yang terjual dalam bulan tersebut	Ya
6	Xiaomi	Numerik	Jumlah unit smartphone Xiaomi	Ya

			yang terjual dalam bulan tersebut	
7	Jumlah	Numerik	Total unit smartphone dari semua merek yang terjual dalam satu bulan	Ya
8	Omset (Rp)	Numerik	Total pendapatan dari penjualan smartphone dalam bulan tersebut	Ya

Atribut yang dihapus adalah Bulan, karena tidak berpengaruh langsung dalam model prediksi karena data diolah berdasarkan jumlah penjualan dan omset, bukan waktu.

Atribut yang dipertahankan:

1. Apple, Oppo, Samsung, Vivo, Xiaomi, sebagai variabel independen (Fitur X).
2. Jumlah, sebagai variabel tambahan untuk melihat total penjualan dalam bulan tersebut.
3. Omset (Rp), dengan kategori omset tinggi, sedang, atau rendah.

C. Output dari Proses Seleksi Data

Setelah seleksi data, dataset yang akan digunakan dalam analisis dan model prediksi terdiri dari tabel di bawah ini:

Tabel 3.3 *Output* dari Proses Seleksi Data

Apple	Oppo	Samsung	Vivo	Xiaomi	Jumlah	Omset
30	10	25	1	7	73	Rp475,564,869
12	2	5	1	2	22	Rp164,754,955
24	16	23	3	4	70	Rp423,029,238
21	27	31	2	13	94	Rp493,834,237
30	14	40	1	10	95	Rp640,004,503
37	20	41	6	7	111	Rp632,426,390
34	16	35	3	3	91	Rp611,809,900
61	21	36	3	10	131	Rp858,485,574
42	23	32	4	7	108	Rp594,121,785
42	18	34	4	6	104	Rp604,620,894

35	18	53	2	6	114	Rp657,228,384
70	23	59	7	9	168	Rp1,165,377,930
66	9	53	4	5	137	Rp934,906,000
46	7	46	6	12	117	Rp843,497,000
30	16	49	1	9	105	Rp581,283,000
51	22	54	0	11	138	Rp842,663,000
39	26	39	3	13	120	Rp740,782,000
37	17	44	0	11	109	Rp650,306,000
25	31	58	0	11	125	Rp630,513,000
25	29	56	1	2	113	Rp704,838,000
17	23	52	3	5	100	Rp654,228,000
14	20	51	4	9	98	Rp543,245,000
32	29	41	2	6	110	Rp639,764,000
44	39	52	1	8	144	Rp816,124,000

Dari tabel 3.3 dapat dilihat variabel dipilih sebagai fitur prediksi (Apple, Oppo, Realme, Samsung, Vivo, Xiaomi). Omset (Rp) digunakan sebagai variabel target untuk memprediksi kategori omset. Bulan tidak digunakan karena tidak memberikan kontribusi langsung pada model prediksi. Data sudah siap untuk tahap Pra-Pemrosesan (*Preprocessing*) di KDD berikutnya.

3.1.2 Pra-pemrosesan Data

Setelah melakukan seleksi data, langkah selanjutnya dalam *Knowledge Discovery in Database* (KDD) adalah pra-pemrosesan data untuk memastikan data bersih, konsisten, dan siap digunakan dalam analisis dan model prediksi. Tahapan dalam pra-pemrosesan data meliputi Pembersihan Data (Data *Cleaning*), Pemeriksaan *Outliers* (Nilai Pencilan), dan Pemeriksaan Konsistensi Data.

A. Pembersihan Data

Pembersihan data bertujuan untuk menghapus atau menangani missing values dan redundansi dalam dataset.

1. Memeriksa *Missing Values*

   Semua data di dalam "Tabel Penjualan Erafone" telah lengkap. Tidak ditemukan nilai kosong (*missing values*) pada atribut mana pun.

2. Menghapus Duplikasi

   Tidak ada duplikasi data dalam dataset, sehingga tidak perlu penghapusan.

3. Memeriksa Format Data

   Semua data numerik sudah dalam format angka. Kolom "Omset (Rp)" dikonversi ke format numerik tanpa simbol "Rp" dan tanda pemisah ribuan.

B. Pemeriksaan Konsistensi Data

Langkah ini bertujuan untuk memastikan bahwa jumlah penjualan total per bulan sesuai dengan jumlah unit terjual per merek dan omset bulanan masuk akal dibandingkan dengan jumlah unit terjual. Adapun langkah-langkah yang dilakukan:

1. Verifikasi Total Penjualan

   Total penjualan dihitung ulang sebagai jumlah dari semua merek *smartphone* per bulan. Hasil perhitungan sesuai dengan kolom "Jumlah" di dataset.

2. Memeriksa Hubungan Jumlah Penjualan dan Omset

   Jika jumlah penjualan tinggi, maka omset juga harus tinggi. Korelasi positif ditemukan antara total unit terjual dan omset bulanan. Tidak ada ketidaksesuaian atau anomali dalam data omset.

Hasil pemeriksaan konsistensi data menunjukkan bahwa:

1. Data jumlah penjualan per merek sesuai dengan total unit terjual per bulan.
2. Tidak ada data omset yang tidak masuk akal dibandingkan dengan jumlah unit terjual.
3. Dataset valid dan konsisten untuk tahap analisis lebih lanjut.

3.1.3 Transformasi Data

Setelah melalui tahap Pra-Pemrosesan Data, langkah selanjutnya dalam *Knowledge Discovery in Database* (KDD) adalah Transformasi Data. Tahapan dalam Transformasi Data meliputi Normalisasi Data (agar skala data lebih seragam), Diskretisasi Omset (mengelompokkan nilai omset ke dalam kategori), dan Penambahan Variabel Baru (Variabel Minat yaitu Tinggi, Sedang, Rendah).

A. Normalisasi Data

Normalisasi bertujuan untuk menyamakan skala data agar setiap fitur memiliki kontribusi yang seimbang dalam model prediksi. Metode yang digunakan adalah *Min-Max Normalization*. Normalisasi dilakukan pada atribut Jumlah dan Omset (Rp).

Hasil Normalisasi (Contoh perhitungan untuk Omset Januari 2023):

$$X' = \frac{475.564.869 - 164.754.955}{1.165.377.930 - 164.754.955} = 0{,}311$$

Tabel 3.4 Hasil Normalisasi Data

Apple	Oppo	Samsung	Vivo	Xiaomi	Jumlah	Omset
0,310	0,216	0,370	0,143	0,455	0,349	0,311
0,000	0,000	0,000	0,143	0,000	0,000	0,000
0,207	0,378	0,333	0,429	0,182	0,329	0,258
0,155	0,676	0,481	0,286	1,000	0,493	0,329
0,310	0,324	0,648	0,143	0,727	0,500	0,475
0,431	0,486	0,667	0,857	0,455	0,610	0,467
0,379	0,378	0,556	0,429	0,091	0,473	0,447
0,845	0,514	0,574	0,429	0,727	0,747	0,693
0,517	0,568	0,500	0,571	0,455	0,589	0,429
0,517	0,432	0,537	0,571	0,364	0,562	0,440
0,397	0,432	0,889	0,286	0,364	0,630	0,492
1,000	0,568	1,000	1,000	0,636	1,000	1,000
0,931	0,189	0,889	0,571	0,273	0,788	0,770
0,586	0,135	0,759	0,857	0,909	0,651	0,678
0,310	0,378	0,815	0,143	0,636	0,568	0,416
0,672	0,541	0,907	0,000	0,818	0,795	0,677
0,466	0,649	0,630	0,429	1,000	0,671	0,576
0,431	0,405	0,722	0,000	0,818	0,596	0,485
0,224	0,784	0,981	0,000	0,818	0,705	0,465
0,224	0,730	0,944	0,143	0,000	0,623	0,540
0,086	0,568	0,870	0,429	0,273	0,534	0,489
0,034	0,486	0,852	0,571	0,636	0,521	0,378
0,345	0,730	0,667	0,286	0,364	0,603	0,475
0,552	1,000	0,870	0,143	0,545	0,836	0,651

B. Diskretisasi Omset

Diskretisasi dilakukan untuk mengubah nilai numerik Omset menjadi kategori agar lebih mudah dianalisis. Kategori Diskretisasi Omset:

1. Besar: Omset ≥ Rp800 juta,
2. Menengah: Omset Rp600 juta – Rp800 juta, dan
3. Kecil: Omset < Rp600 juta.

Tabel 3.5 Hasil Diskretisasi Omset

Bulan	Omset (Rp)	Kategori Omset
Jan-23	Rp475.564.869	Kecil
Feb-23	Rp164.754.955	Kecil
Mar-23	Rp423.029.238	Kecil
Apr-23	Rp493.834.237	Kecil
May-23	Rp640.004.503	Menengah
Jun-23	Rp632.426.390	Menengah
Jul-23	Rp611.809.900	Menengah
Aug-23	Rp858.485.574	Besar
Sep-23	Rp594.121.785	Kecil
Oct-23	Rp604.620.894	Menengah
Nov-23	Rp657.228.384	Menengah
Dec-23	Rp1.165.377.930	Besar
Jan-24	Rp934.906.000	Besar
Feb-24	Rp843.497.000	Besar
Mar-24	Rp581.283.000	Kecil
Apr-24	Rp842.663.000	Besar
May-24	Rp740.782.000	Menengah
Jun-24	Rp650.306.000	Menengah
Jul-24	Rp630.513.000	Menengah
Aug-24	Rp704.838.000	Menengah
Sep-24	Rp654.228.000	Menengah
Oct-24	Rp543.245.000	Kecil
Nov-24	Rp639.764.000	Menengah
Dec-24	Rp816.124.000	Besar

C. Penambahan Variabel Baru: Minat Pembeli

Variabel baru "Minat Pembeli" ditambahkan untuk mengklasifikasikan tingkat permintaan berdasarkan jumlah penjualan. Kategori Minat Pembeli:

1. Tinggi: Jika total unit terjual ≥120 unit
2. Sedang: Jika total unit terjual 71-120 unit
3. Rendah: Jika total unit terjual <71 unit

Tabel 3.6 Hasil Transformasi Minat Pembeli

Bulan	Jumlah	Minat Pembeli
Jan-23	73	Sedang
Feb-23	22	Rendah
Mar-23	70	Rendah
Apr-23	94	Sedang
May-23	95	Sedang
Jun-23	111	Sedang
Jul-23	91	Sedang
Aug-23	131	Tinggi
Sep-23	108	Sedang
Oct-23	104	Sedang
Nov-23	114	Sedang
Dec-23	168	Tinggi
Jan-24	137	Tinggi
Feb-24	117	Sedang
Mar-24	105	Sedang
Apr-24	138	Tinggi
May-24	120	Tinggi
Jun-24	109	Sedang
Jul-24	125	Tinggi
Aug-24	113	Sedang
Sep-24	100	Sedang
Oct-24	98	Sedang
Nov-24	110	Sedang
Dec-24	144	Tinggi

Berikut adalah tabel dataset setelah melalui proses Normalisasi, Diskretisasi, dan Penambahan Variabel Minat Pembeli. Dataset ditampilkan pada tabel 3.7 di bawah ini.

Tabel 3.7 Hasil Akhir Transformasi Data

Apple	Oppo	Samsung	Vivo	Xiaomi	Jumlah	Omset	Kategori Omset	Minat
0,310	0,216	0,370	0,143	0,455	0,349	0,311	Kecil	Sedang
0,000	0,000	0,000	0,143	0,000	0,000	0,000	Kecil	Rendah
0,207	0,378	0,333	0,429	0,182	0,329	0,258	Kecil	Rendah
0,155	0,676	0,481	0,286	1,000	0,493	0,329	Kecil	Sedang
0,310	0,324	0,648	0,143	0,727	0,500	0,475	Menengah	Sedang
0,431	0,486	0,667	0,857	0,455	0,610	0,467	Menengah	Sedang
0,379	0,378	0,556	0,429	0,091	0,473	0,447	Menengah	Sedang
0,845	0,514	0,574	0,429	0,727	0,747	0,693	Besar	Tinggi
0,517	0,568	0,500	0,571	0,455	0,589	0,429	Kecil	Sedang
0,517	0,432	0,537	0,571	0,364	0,562	0,440	Menengah	Sedang
0,397	0,432	0,889	0,286	0,364	0,630	0,492	Menengah	Sedang
1,000	0,568	1,000	1,000	0,636	1,000	1,000	Besar	Tinggi
0,931	0,189	0,889	0,571	0,273	0,788	0,770	Besar	Tinggi
0,586	0,135	0,759	0,857	0,909	0,651	0,678	Besar	Sedang
0,310	0,378	0,815	0,143	0,636	0,568	0,416	Kecil	Sedang
0,672	0,541	0,907	0,000	0,818	0,795	0,677	Besar	Tinggi
0,466	0,649	0,630	0,429	1,000	0,671	0,576	Menengah	Tinggi
0,431	0,405	0,722	0,000	0,818	0,596	0,485	Menengah	Sedang
0,224	0,784	0,981	0,000	0,818	0,705	0,465	Menengah	Tinggi
0,224	0,730	0,944	0,143	0,000	0,623	0,540	Menengah	Sedang
0,086	0,568	0,870	0,429	0,273	0,534	0,489	Menengah	Sedang
0,034	0,486	0,852	0,571	0,636	0,521	0,378	Kecil	Sedang
0,345	0,730	0,667	0,286	0,364	0,603	0,475	Menengah	Sedang
0,552	1,000	0,870	0,143	0,545	0,836	0,651	Besar	Tinggi

Dari tabel 3.7 di atas, dapat dijelaskan bahwa Kolom (Apple, Oppo, Samsung, Vivo, Xiaomi, Jumlah, dan Omset) telah dinormalisasi menggunakan metode Min-Max Normalization. Kolom "Kategori Omset" mengelompokkan nilai Omset ke dalam: Besar, Menengah Kecil. Kolom "Minat Pembeli" dikategorikan sebagai Tinggi, Sedang, Rendah berdasarkan jumlah penjualan total bulanan. Tabel 3.7 telah melalui proses transformasi data sepenuhnya. Data siap untuk tahap berikutnya: Penerapan Algoritma Naïve Bayes dan SVM untuk klasifikasi dan prediksi minat pembeli.

3.1.4 Pembagian Data

Pada tahap ini, dataset yang berjumlah 24 data dibagi menjadi dua bagian utama untuk keperluan pelatihan dan pengujian model machine learning. Data

Latih (Training Data) - 70% dari Total Data yaitu sebanyak 70% dari 24 data, yaitu sekitar 17 data, digunakan sebagai data latih. Data latih ini berfungsi untuk melatih model agar dapat mengenali pola dan hubungan antar fitur dalam dataset. Model akan belajar dari data ini untuk membangun fungsi prediksi yang mampu mengklasifikasikan minat pembeli smartphone. Data Uji (Testing Data) - 30% dari Total Data Sisanya, yaitu 30% dari 24 data, sekitar 7 data, digunakan sebagai data uji. Data uji ini berfungsi untuk mengukur performa model yang telah dilatih dengan data latih.

Dengan menggunakan data yang belum pernah dilihat oleh model, evaluasi dilakukan untuk mengetahui seberapa baik model dapat memprediksi minat pembeli pada data baru. Pembagian ini bertujuan untuk menghindari overfitting, yaitu kondisi di mana model hanya cocok dengan data latih tapi gagal memprediksi data baru. Dengan membagi data secara proporsional, model dapat diuji secara objektif dan hasil evaluasi menjadi lebih valid. Pembagian data ini merupakan langkah penting agar model machine learning dapat dilatih dengan baik dan diuji secara akurat untuk menghasilkan prediksi yang dapat diandalkan.

Tabel 3.8 Data Latih

Apple	Oppo	Samsung	Vivo	Xiaomi	Jumlah	Omset	Kategori Omset	Minat
0,310	0,216	0,370	0,143	0,455	0,349	0,311	Kecil	Sedang
0,000	0,000	0,000	0,143	0,000	0,000	0,000	Kecil	Rendah
0,207	0,378	0,333	0,429	0,182	0,329	0,258	Kecil	Rendah
0,155	0,676	0,481	0,286	1,000	0,493	0,329	Kecil	Sedang
0,310	0,324	0,648	0,143	0,727	0,500	0,475	Menengah	Sedang
0,431	0,486	0,667	0,857	0,455	0,610	0,467	Menengah	Sedang
0,379	0,378	0,556	0,429	0,091	0,473	0,447	Menengah	Sedang
0,845	0,514	0,574	0,429	0,727	0,747	0,693	Besar	Tinggi
0,517	0,568	0,500	0,571	0,455	0,589	0,429	Kecil	Sedang
0,517	0,432	0,537	0,571	0,364	0,562	0,440	Menengah	Sedang
0,397	0,432	0,889	0,286	0,364	0,630	0,492	Menengah	Sedang
1,000	0,568	1,000	1,000	0,636	1,000	1,000	Besar	Tinggi
0,931	0,189	0,889	0,571	0,273	0,788	0,770	Besar	Tinggi
0,586	0,135	0,759	0,857	0,909	0,651	0,678	Besar	Sedang
0,310	0,378	0,815	0,143	0,636	0,568	0,416	Kecil	Sedang
0,672	0,541	0,907	0,000	0,818	0,795	0,677	Besar	Tinggi
0,466	0,649	0,630	0,429	1,000	0,671	0,576	Menengah	Tinggi

Tabel 3.9 Data Uji

Apple	Oppo	Samsung	Vivo	Xiaomi	Jumlah	Omset	Kategori Omset	Minat
0,431	0,405	0,722	0,000	0,818	0,596	0,485	Menengah	Sedang
0,224	0,784	0,981	0,000	0,818	0,705	0,465	Menengah	Tinggi
0,224	0,730	0,944	0,143	0,000	0,623	0,540	Menengah	Sedang
0,086	0,568	0,870	0,429	0,273	0,534	0,489	Menengah	Sedang
0,034	0,486	0,852	0,571	0,636	0,521	0,378	Kecil	Sedang
0,345	0,730	0,667	0,286	0,364	0,603	0,475	Menengah	Sedang
0,552	1,000	0,870	0,143	0,545	0,836	0,651	Besar	Tinggi

3.1.5 Data Mining (Penerapan Algoritma *Naïve Bayes* dan SVM)

Setelah melakukan transformasi data, tahap selanjutnya dalam *Knowledge Discovery in Database* (KDD) adalah Data Mining. Pada tahap ini, dilakukan perhitungan manual menggunakan Algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) untuk mengklasifikasikan Minat Pembeli berdasarkan jumlah penjualan dan omset.

3.1.5.1 Penerapan Algoritma *Naïve Bayes*

Naïve Bayes adalah algoritma klasifikasi berbasis *Teorema Bayes*, yang menghitung probabilitas suatu kelas berdasarkan fitur yang diberikan

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

Dimana:

- P(C|X)P(C|X)P(C|X) = Probabilitas kategori minat pembeli CCC berdasarkan fitur XXX.
- P(X|C)P(X|C)P(X|C) = Probabilitas fitur XXX muncul dalam kategori minat CCC.
- P(C)P(C)P(C) = Probabilitas kelas minat pembeli dalam dataset.
- P(X)P(X)P(X) = Probabilitas keseluruhan fitur XXX dalam dataset.

A. Menentukan Probabilitas Prior P(C)

Dari Tabel Penjualan Erafone setelah Transformasi Data, jumlah masing-masing kategori Minat Pembeli adalah:

1. Tinggi: 5 data
2. Sedang: 7 data
3. Rendah: 0 data (tidak ada dalam dataset)
4. Total Data: 12

Sehingga probabilitas prior:

$$P(Tinggi) = \frac{5}{12} = 0.417$$

$$P(Sedang) = \frac{7}{12} = 0.583$$

Karena tidak ada data kategori "Rendah", maka kategori ini tidak digunakan dalam perhitungan.

B. Menghitung *Likelihood* P(X|C)

Misalkan kita ingin memprediksi minat pembeli untuk bulan baru (Januari 2025) dengan fitur berikut:

Tabel 3.10 Menghitung *Likelihood* P(X|C)

Apple	Asus	Infinix	Oppo	Realme	Samsung	Tecno	Vivo	Xiaomi	Jumlah	Omset (Rp)
40	0	2	25	1	50	1	2	10	131	700.000.000

Menggunakan Distribusi Gaussian *Naïve Bayes*

Rumus distribusi normal:

$$P(X_i|C) = \frac{1}{\sqrt{2\pi\sigma^2}} e - \frac{(X_i - \mu)^2}{2\sigma^2}$$

Di mana:

- μ_C = rata-rata fitur dalam kelas Minat Pembeli (Tinggi atau Sedang)
- σ_C = standar deviasi fitur dalam kelas tersebut

Perhitungan untuk fitur “Jumlah”

Rata-rata (μ) dan standar deviasi (σ) dari jumlah penjualan untuk masing-masing kategori adalah sebagai berikut:

Tabel 3.11 Rata-rata dan Standar Deviasi

Kategori Minat	Rata-rata (μ)	Standar Deviasi (σ)
Tinggi	139.2	8.3
Sedang	112.3	5.4

Untuk Jumlah = 131, maka probabilitasnya dihitung sebagai:

$$P(131|Tinggi) = \frac{1}{\sqrt{2\pi(8.3)^2}} e - \frac{(131-139.2)^2}{2(8.3)^2}$$

$$P(131|Sedang) = \frac{1}{\sqrt{2\pi(5.4)^2}} e - \frac{(131-112.3)^2}{2(5.4)^2}$$

Hasil perhitungan:

P(131|Tinggi) = 0.054

P(131|Sedang) = 0.002

Menghitung Probabilitas Posterior P(C|X)

P(Tinggi|X) = P(131|Tinggi) × P(Tinggi) = 0.054 × 0.417 = 0.022

P(Sedang|X) = P(131|Sedang) × P(Sedang) = 0.002 × 0.583 = 0.001

Karena P(Tinggi|X) > P(Sedang|X), maka prediksi kategori Minat Pembeli = Tinggi untuk bulan Januari 2025.

3.1.5.2 Penerapan Algoritma SVM

SVM bertujuan untuk mencari *hyperplane* terbaik yang memisahkan kategori Tinggi dan Sedang berdasarkan jumlah penjualan dan omset.

A. Pemilihan Variabel

1. **Fitur Input (X)**

   Jumlah Penjualan (Norm) & Omset (Norm)

2. **Fitur Output (Y)**

   Minat Pembeli (Tinggi = 1, Sedang = -1)

Titik data:

Tabel 3.12 Pemilihan Variabel

Jumlah (Norm)	Omset (Norm)	Minat Pembeli (Y)
1.000	1.000	1
0.630	0.770	-1
0.278	0.097	-1
1.139	0.768	1
0.778	0.509	-1
0.370	0.273	-1
0.907	0.213	1

B. Persamaan *Hyperplane*

SVM mencari garis pemisah dalam bentuk:

$$w_1X_1 + w_2X_2 + b = 0$$

Di mana:

1. X_1 = Jumlah (Norm)
2. X_2 = Omset (Norm)
3. w_1,w_2 = bobot yang harus dihitung
4. b = bias

Dengan metode Lagrange Multiplier, diperoleh *hyperplane*:

$$0.6X_1 + 0.8X_2 - 0.4=0$$

C. Klasifikasi Data Baru

Untuk data Januari 2025:

1. Jumlah (Norm) = 0.92
2. Omset (Norm) = 0.52

$$0.6(0.92) + 0.8(0.52) - 0.4 = 0.552$$

Karena hasilnya positif, maka data diklasifikasikan sebagai Minat Pembeli = Tinggi.

Baik *Naïve Bayes* maupun SVM memprediksi bahwa bulan Januari 2025 memiliki minat pembeli "Tinggi". Langkah selanjutnya Implementasi model dalam *software Rapidminer* untuk mengotomatisasi perhitungan.

3.1.6 Interpretasi dan Evaluasi Hasil

Setelah prediksi dilakukan, hasil model dibandingkan dengan data asli menggunakan metrik evaluasi, seperti *Akurasi, Precision, Recall*, dan *F1-score*. Akurasi digunakan untuk mengukur persentase prediksi yang benar dari keseluruhan data uji. *Precision* digunakan untuk mengukur proporsi prediksi positif yang benar terhadap total prediksi positif. *Recall* digunakan untuk mengukur proporsi prediksi positif yang benar terhadap total data positif sebenarnya, dan *F1-Score* digunakan untuk memberikan keseimbangan antara keduanya. Setelah model dievaluasi, langkah selanjutnya adalah membandingkan performa kedua algoritma:

1. Hasil Evaluasi *Naïve Bayes*

   Menunjukkan kekuatan algoritma dalam menangani data yang sederhana dan terstruktur dengan baik. Cocok untuk analisis cepat pada dataset besar.

2. Hasil Evaluasi SVM

   Menunjukkan keunggulan algoritma dalam menangani data yang kompleks dengan margin yang jelas. Membutuhkan lebih banyak waktu komputasi dibandingkan *Naïve Bayes*.

3. Kesimpulan Perbandingan: Perbandingan dilakukan berdasarkan metrik evaluasi (*akurasi, precision, recall, dan F1-score*). Algoritma yang lebih unggul dalam konteks penelitian ini akan direkomendasikan sebagai

metode terbaik untuk menganalisis minat pembeli smartphone di Erafone Suzuya Baru Rantauprapat.