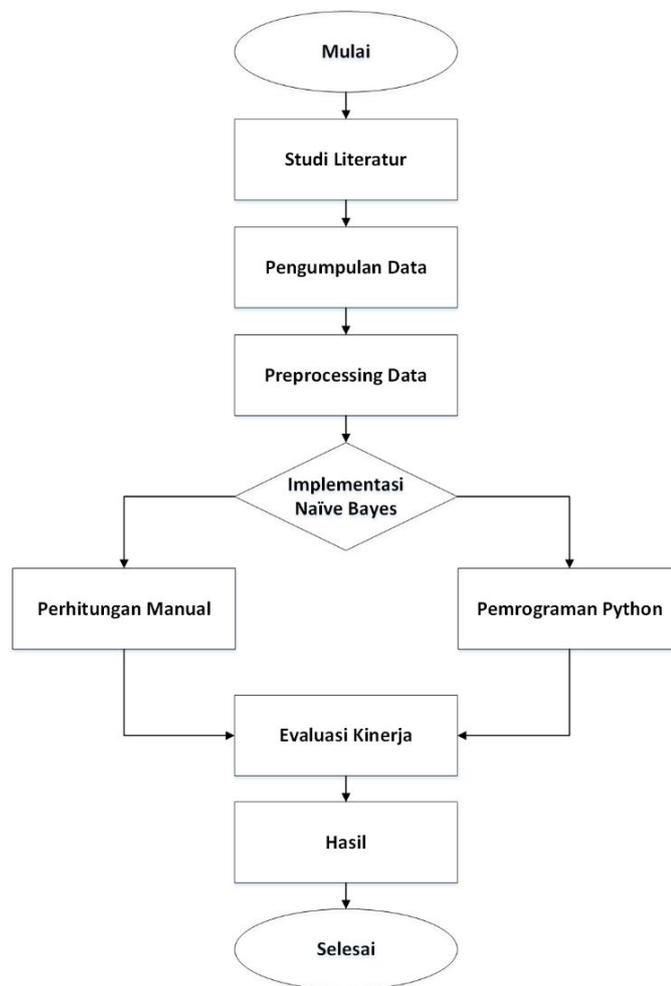


BAB III

ANALISIS DAN PERANCANGAN

Penelitian ini menggunakan pendekatan kuantitatif untuk menganalisis dan memprediksi kedisiplinan karyawan PT. PP London Sumatra Tbk, Sei Rumbiya Estate. Algoritma *Naïve Bayes* digunakan sebagai metode klasifikasi pada penelitian ini. Diagram alur kerja dari penelitian ini diperlihatkan pada Gambar 3.1 di bawah ini.



Gambar 3. 1 Alur Penelitian

2.6 Studi Literatur

Studi literatur merupakan tahap awal dalam penelitian yang bertujuan untuk memperoleh pemahaman mendalam terkait topik yang diteliti. Tahapan ini dilakukan dengan mengumpulkan dan menganalisis berbagai sumber referensi yang relevan, seperti jurnal ilmiah, buku akademik, konferensi penelitian, dan laporan penelitian terdahulu. Sumber referensi yang digunakan dalam studi literatur berasal dari database akademik seperti Google Scholar, IEEE Xplore, ScienceDirect, dan jurnal nasional serta internasional yang relevan dengan bidang penelitian ini. Studi literatur dalam penelitian ini mencakup beberapa aspek utama:

1. Konsep Machine Learning

Dalam penelitian ini, *machine learning* digunakan untuk mengidentifikasi pola dalam kedisiplinan karyawan. Referensi yang dikaji meliputi teknik dan metode machine learning yang umum digunakan dalam analisis kedisiplinan karyawan.

2. Algoritma *Naïve Bayes*

Studi literatur mencakup prinsip kerja *Naïve Bayes*, keunggulan dan keterbatasannya, serta implementasinya dalam berbagai kasus klasifikasi.

3. Faktor-Faktor yang Mempengaruhi Kedisiplinan Karyawan

Faktor internal seperti kesehatan karyawan, motivasi kerja, dan kepuasan kerja. Faktor eksternal seperti kondisi cuaca, jarak tempat tinggal ke tempat kerja, dan kebijakan perusahaan. Studi literatur dilakukan untuk memahami faktor-faktor ini dan menentukan variabel yang akan digunakan dalam penelitian.

4. Studi Terdahulu yang Relevan

Menganalisis penelitian sebelumnya yang menggunakan metode *Naive Bayes* dalam prediksi kedisiplinan karyawan atau bidang lain yang serupa. Meninjau hasil penelitian sebelumnya untuk mengidentifikasi pendekatan yang efektif dan potensi perbaikan dalam penelitian ini.

2.7 Pengumpulan Data

Pengumpulan data adalah langkah penting dalam penelitian yang bertujuan untuk mengumpulkan informasi yang diperlukan untuk analisis dan pengembangan model. Berikut adalah penjelasan lebih detail dan komprehensif mengenai proses ini, termasuk sumber data dan variabel yang digunakan. Proses pengumpulan data dilakukan dengan beberapa tahapan sebagai berikut:

A. Sumber Data

Data kedisiplinan karyawan diambil dari sistem absensi berbasis fingerprint yang digunakan oleh perusahaan. Sistem ini mencatat waktu masuk dan keluar karyawan secara otomatis. Mengakses database internal perusahaan yang menyimpan informasi terkait karyawan, seperti riwayat kehadiran, cuti, dan lembur. Mengumpulkan dokumen terkait kebijakan kehadiran, laporan kinerja, dan catatan lainnya yang dapat memberikan konteks tambahan untuk analisis.

B. Variabel yang Digunakan

Variabel yang digunakan dalam penelitian ini dapat dibagi menjadi dua kategori: variabel input (*independent variables*) dan variabel target (*dependent variable*).

1. Variabel Input

Variabel input adalah fitur-fitur yang digunakan untuk memprediksi variabel target. Dalam konteks penelitian ini, variabel input meliputi:

- Jumlah Hari Sakit (S1-S5): Menghitung jumlah hari sakit yang diambil oleh karyawan dalam periode tertentu. Ini dapat mencakup kategori seperti sakit ringan, sedang, dan berat.
- Jumlah Lembur: Total jam lembur yang dilakukan oleh karyawan. Lembur dapat menjadi indikator komitmen dan beban kerja.
- Total Hari Kerja: Jumlah total hari kerja dalam periode yang dianalisis. Ini memberikan konteks untuk menghitung persentase kedisiplinan

2. Variabel Target

Variabel target adalah hasil yang ingin diprediksi oleh model. Dalam penelitian ini, variabel target adalah: Kriteria yang menunjukkan apakah karyawan dianggap "Disiplin" atau "Tidak Disiplin" berdasarkan analisis kedisiplinan mereka.

Tabel 3. 1 Pengumpulan Dataset

No	Nama Karyawan	S1	S2	S3	S4	S5	Lembur	Total Kerja	Kriteria
1	Riyanti	10	2	0	0	0	0	17	Tidak Disiplin
2	Suparno	0	0	3	0	0	6	32	Disiplin
3	Junaidi	0	0	0	0	0	7	36	Disiplin
4	Lamhot Tambunan	2	1	0	0	0	0	26	Tidak Disiplin
5	Supiyat	0	0	0	1	0	0	28	Tidak Disiplin
6	Muhammad Sapparudin	0	0	0	0	1	0	28	Tidak Disiplin
7	Sumawardi Purba	0	0	0	0	1	0	28	Tidak Disiplin
8	Muhammad Tohali Syahputra	8	3	0	0	0	0	18	Tidak Disiplin
9	Sunarto N	0	0	0	0	0	0	29	Disiplin
10	Legiman	0	0	0	0	1	0	28	Tidak Disiplin
11	Suparno	0	0	0	0	0	4	33	Disiplin
12	Sudarwan	0	0	0	0	0	0	29	Disiplin
13	Kusmanto	0	0	0	0	0	7	36	Disiplin
14	Efan Agus Pranata	4	0	0	0	0	0	25	Tidak Disiplin
15	Ngadeliman	0	0	0	0	0	0	29	Disiplin
16	Yus Efendi	0	0	0	1	0	0	28	Tidak Disiplin
17	Andrianto Syahputra	1	0	0	0	0	7	35	Disiplin
18	Surati	0	0	1	0	0	1	29	Disiplin
19	Sunardi	4	0	0	0	0	0	25	Tidak Disiplin
20	Indra Bakti	2	1	0	0	0	7	33	Disiplin

Keterangan :

Tabel yang memiliki nilai atau angka akan diberi nilai 1 dan yang tidak memiliki nilai diberi nilai 0.

Tabel 3. 2 Dataset Penelitian

No	Nama Karyawan	S1	S2	S3	S4	S5	Lembur	Kriteria
1	Riyanti	1	1	0	0	0	0	Tidak Disiplin
2	Suparno	0	0	1	0	0	6	Disiplin
3	Junaidi	0	0	0	0	0	7	Disiplin
4	Lamhot Tambunan	1	1	0	0	0	0	Tidak Disiplin
5	Supiyat	0	0	0	1	0	0	Tidak Disiplin
6	Muhammad Sapparudin	0	0	0	0	1	0	Tidak Disiplin
7	Sumawardi Purba	0	0	0	0	1	0	Tidak Disiplin
8	Muhammad Tohali Syahputra	1	1	0	0	0	0	Tidak Disiplin
9	Sunarto N	0	0	0	0	0	0	Disiplin
10	Legiman	0	0	0	0	1	0	Tidak Disiplin
11	Suparno	0	0	0	0	0	4	Disiplin
12	Sudarwan	0	0	0	0	0	0	Disiplin
13	Kusmanto	0	0	0	0	0	7	Disiplin
14	Efan Agus Pranata	1	0	0	0	0	0	Tidak Disiplin
15	Ngadeliman	0	0	0	0	0	0	Disiplin
16	Yus Efendi	0	0	0	1	0	0	Tidak Disiplin
17	Andrianto Syahputra	1	0	0	0	0	7	Disiplin
18	Surati	0	0	1	0	0	0	Tidak Disiplin
19	Sunardi	1	0	0	0	0	0	Tidak Disiplin
20	Indra Bakti	1	1	0	0	0	7	Disiplin

Keterangan:

- **S1 (Sakit):** Jumlah hari karyawan tidak hadir karena sakit biasa. Ini mencakup absensi yang tidak memerlukan perawatan lebih lanjut.
- **S2 (Rujukan):** Jumlah hari karyawan tidak hadir karena dirujuk ke fasilitas kesehatan lain. Ini menunjukkan bahwa karyawan memerlukan perawatan lebih lanjut.
- **S3 (Sakit di Rumah Sakit):** Jumlah hari karyawan tidak hadir karena dirawat di rumah sakit. Ini menunjukkan tingkat keparahan kondisi kesehatan.
- **S4 (Sakit di Rumah Sakit 4 Bulan):** Jumlah hari karyawan tidak hadir karena sakit yang memerlukan perawatan di rumah sakit selama 4 bulan. Ini menunjukkan masalah kesehatan yang serius dan berkepanjangan.
- **S5 (Sakit Selama 5 Bulan):** Jumlah hari karyawan tidak hadir karena sakit yang memerlukan perawatan di rumah sakit selama 5 bulan. Ini juga menunjukkan masalah kesehatan yang sangat serius dan dapat mempengaruhi produktivitas jangka panjang.
- **Lembur:** Total jam lembur yang dilakukan oleh karyawan. Lembur dapat menjadi indikator komitmen dan beban kerja.
- **Total Kerja** = (Jumlah Hari Kerja yaitu 29 hari + Lembur) – (S1+S2+S3+S4+S5).
- **Kriteria:** Karyawan dikategorikan sebagai " Disiplin " jika jumlah total kerja mereka dalam sebulan mencapai atau melebihi 29 hari. Karyawan dikategorikan sebagai "Tidak Disiplin " jika jumlah total kerja mereka kurang dari 29 hari.

Dengan kriteria yang jelas untuk mengkategorikan karyawan sebagai Disiplin atau tidak Disiplin, peneliti dapat lebih efektif dalam menganalisis data kedisiplinan. Kriteria ini membantu dalam mengidentifikasi pola kedisiplinan dan dapat digunakan untuk merumuskan strategi manajemen yang lebih baik untuk meningkatkan kedisiplinan dan produktivitas karyawan. Dataset ini memberikan gambaran yang jelas tentang kedisiplinan karyawan dan faktor-faktor yang mempengaruhi kinerja mereka. Dengan menggunakan data ini, peneliti dapat melakukan analisis lebih lanjut untuk memprediksi kedisiplinan karyawan dan mengidentifikasi pola yang mungkin ada dalam data. Ini juga dapat membantu dalam merumuskan strategi untuk meningkatkan kedisiplinan dan produktivitas karyawan di perusahaan.

2.8 Preprocessing Data

Preprocessing data adalah langkah penting dalam analisis data yang bertujuan untuk mempersiapkan data agar siap digunakan dalam model analisis atau pembelajaran mesin. Berikut adalah langkah-langkah untuk *preprocessing data*:

1. *Cleaning Data* (Pembersihan Data)

- Semua kolom dalam dataset (S1, S2, S3, S4, S5, Lembur, Kriteria) dipertahankan karena semuanya relevan untuk analisis kedisiplinan karyawan.
- Kolom Total Kerja dihapus karena hanya digunakan sebagai indikator penentuan kriteria “Disiplin” atau “Tidak Disiplin”.

2. *Handling Missing Values* (Menangani Data Hilang)

- Dalam dataset ini, tidak ada nilai hilang yang terdeteksi. Semua kolom memiliki data yang lengkap.

3. Normalisasi Data

- Normalisasi diterapkan pada variabel numerik (S1, S2, S3, S4, S5, dan Lembur) untuk memastikan bahwa semua fitur berada dalam skala yang sama. Hasilnya diperlihatkan pada Tabel 3.3.

Tabel 3. 3 Dataset Hasil Normalisasi

No	Nama Karyawan	S1	S2	S3	S4	S5	Lembur	Kriteria
1	Riyanti	1	1	0	0	0	0,00	Tidak Disiplin
2	Suparno	0	0	1	0	0	0,86	Disiplin
3	Junaidi	0	0	0	0	0	1,00	Disiplin
4	Lamhot Tambunan	1	1	0	0	0	0,00	Tidak Disiplin
5	Supiyat	0	0	0	1	0	0,00	Tidak Disiplin
6	Muhammad Sapparudin	0	0	0	0	1	0,00	Tidak Disiplin
7	Sumawardi Purba	0	0	0	0	1	0,00	Tidak Disiplin
8	Muhammad Tohali Syahputra	1	1	0	0	0	0,00	Tidak Disiplin
9	Sunarto N	0	0	0	0	0	0,00	Disiplin
10	Legiman	0	0	0	0	1	0,00	Tidak Disiplin
11	Suparno	0	0	0	0	0	0,57	Disiplin
12	Sudarwan	0	0	0	0	0	0,00	Disiplin
13	Kusmanto	0	0	0	0	0	1,00	Disiplin
14	Efan Agus Pranata	1	0	0	0	0	0,00	Tidak Disiplin
15	Ngadeliman	0	0	0	0	0	0,00	Disiplin
16	Yus Efendi	0	0	0	1	0	0,00	Tidak Disiplin
17	Andrianto Syahputra	1	0	0	0	0	1,00	Disiplin
18	Surati	0	0	1	0	0	0,00	Tidak Disiplin
19	Sunardi	1	0	0	0	0	0,00	Tidak Disiplin
20	Indra Bakti	1	1	0	0	0	1,00	Disiplin

4. Pembagian Data

Dataset pada tabel 3.2 selanjutnya dilakukan proses pembagian data dengan rasio 80% untuk data latih (*training data*) dan 20% untuk data uji (*testing data*). Hasilnya diperlihatkan pada tabel 3.3 dan tabel 3.4.

Tabel 3. 4 Data Latih (80%)

No	Nama Karyawan	S1	S2	S3	S4	S5	Lembur	Kriteria
1	Riyanti	1	1	0	0	0	0,00	Tidak Disiplin
2	Suparno	0	0	1	0	0	0,86	Disiplin
3	Junaidi	0	0	0	0	0	1,00	Disiplin
4	Lamhot Tambunan	1	1	0	0	0	0,00	Tidak Disiplin
5	Supiyat	0	0	0	1	0	0,00	Tidak Disiplin
6	Muhammad Sapparudin	0	0	0	0	1	0,00	Tidak Disiplin
7	Sumawardi Purba	0	0	0	0	1	0,00	Tidak Disiplin
8	Muhammad Tohali Syahputra	1	1	0	0	0	0,00	Tidak Disiplin
9	Sunarto N	0	0	0	0	0	0,00	Disiplin
10	Legiman	0	0	0	0	1	0,00	Tidak Disiplin
11	Suparno	0	0	0	0	0	0,57	Disiplin
12	Sudarwan	0	0	0	0	0	0,00	Disiplin
13	Kusmanto	0	0	0	0	0	1,00	Disiplin
14	Efan Agus Pranata	1	0	0	0	0	0,00	Tidak Disiplin
15	Ngadeliman	0	0	0	0	0	0,00	Disiplin
16	Yus Efendi	0	0	0	1	0	0,00	Tidak Disiplin

Tabel 3. 5 Data Uji (20%)

17	Andrianto Syahputra	1	0	0	0	0	1,00	Disiplin
18	Surati	0	0	1	0	0	0,00	Tidak Disiplin
19	Sunardi	1	0	0	0	0	0,00	Tidak Disiplin
20	Indra Bakti	1	1	0	0	0	1,00	Disiplin

Tabel 3.4 berisi Data Latih yang terdiri dari 16 karyawan yang dipilih dari Tabel 3.3 Dataset. Data ini mencakup informasi penting mengenai kedisiplinan karyawan, termasuk jumlah hari sakit dalam berbagai kategori (S1 hingga S5), jumlah lembur, total hari kerja, dan kriteria kedisiplinan (Disiplin atau Tidak Disiplin). Data latih digunakan untuk melatih model analisis atau pembelajaran mesin. Dengan menggunakan data ini, model dapat belajar dari pola yang ada, seperti hubungan antara jumlah hari sakit dan kriteria kedisiplinan karyawan. Data latih mencakup karyawan dengan berbagai tingkat kedisiplinan, sehingga model dapat memahami variasi dalam data dan membuat prediksi yang lebih akurat.

Tabel 3.5 berisi Data Uji yang terdiri dari 4 karyawan yang juga diambil dari Tabel 3.3 Dataset. Data ini memiliki struktur yang sama dengan data latih, mencakup informasi tentang jumlah hari sakit, lembur, total hari kerja, dan kriteria kedisiplinan. Data uji digunakan untuk menguji kinerja model yang telah dilatih. Dengan menggunakan data ini, peneliti dapat mengevaluasi seberapa baik model dapat memprediksi kriteria kedisiplinan berdasarkan data yang belum pernah dilihat sebelumnya. Data uji penting untuk memastikan bahwa model tidak hanya menghafal data latih, tetapi juga dapat generalisasi dan memberikan prediksi yang akurat pada data baru.

Secara keseluruhan, pembagian data menjadi data latih dan data uji adalah langkah penting dalam proses analisis dan pengembangan model. Data latih memberikan informasi yang diperlukan untuk melatih model, sementara data uji memungkinkan peneliti untuk mengevaluasi kinerja model dan memastikan bahwa model tersebut dapat diterapkan secara efektif dalam situasi nyata. Dengan cara ini, penelitian dapat memberikan wawasan yang lebih baik tentang kedisiplinan karyawan dan faktor-faktor yang mempengaruhinya.

2.9 Implementasi *Naïve Bayes*

Setelah proses preprocessing selesai, algoritma *Naïve Bayes* diterapkan dengan langkah-langkah sebagai berikut:

1. Menghitung Probabilitas Prior ($P(\text{Kriteria})$): Probabilitas setiap kelas dalam variabel target Kriteria.

Dari Data Latih (80%), kita menghitung probabilitas setiap kelas kedisiplinan:

Tabel 3. 6 Kriteria

Kriteria	Jumlah Kemunculan	$P(\text{Kriteria})$
Disiplin	7	$7/16 = 0,44$
Tidak Disiplin	9	$9/16 = 0,56$

Jadi, masing-masing kelas Kriteria memiliki probabilitas prior yang berbeda, yaitu 0,44 untuk kelas “Disiplin”, dan 0,56 untuk kelas “Tidak Disiplin”.

2. Menghitung Probabilitas Kondisional ($P(\text{Fitur} \mid \text{Kriteria})$)

Kita akan menghitung distribusi masing-masing fitur untuk setiap kategori Kriteria (Disiplin, Tidak Disiplin).

a) Probabilitas Kondisional $P(S1 \mid \text{Kriteria})$ **Tabel 3. 7** Kriteria S1

S1	Disiplin	Tidak Disiplin
0	$7/7 = 1$	$5/9 = 0,56$
1	$0/7 = 0$	$4/9 = 0,44$

b) Probabilitas Kondisional $P(S2 \mid \text{Kriteria})$ **Tabel 3. 8** Kriteria S2

S2	Disiplin	Tidak Disiplin
0	$7/7 = 1$	$6/9 = 0,67$
1	$0/7 = 0$	$3/9 = 0,33$

c) Probabilitas Kondisional $P(S3 \mid \text{Kriteria})$ **Tabel 3. 9** Kriteria S3

S3	Disiplin	Tidak Disiplin
0	$6/7 = 0,86$	$9/9 = 1$
1	$1/7 = 0,14$	$0/9 = 0$

d) Probabilitas Kondisional $P(S4 \mid \text{Kriteria})$ **Tabel 3. 10** Kriteria S4

S4	Disiplin	Tidak Disiplin
0	$7/7 = 1$	$7/9 = 0,78$
1	$0/7 = 0$	$2/9 = 0,22$

e) Probabilitas Kondisional $P(S5 | \text{Kriteria})$

Tabel 3. 11 Kriteria S5

S5	Disiplin	Tidak Disiplin
0	$7/7 = 1$	$5/9 = 0,56$
1	$0/7 = 0$	$4/9 = 0,44$

f) Probabilitas Kondisional $P(\text{Lembur} | \text{Kriteria})$

Tabel 3. 12 Kriteria Lembur

S4	Disiplin	Tidak Disiplin
0	$3/7 = 0,43$	$9/9 = 1$
0,57	$1/7 = 0,14$	$0/9 = 0$
0,86	$1/7 = 0,14$	$0/9 = 0$
1	$2/7 = 0,29$	$0/9 = 0$

3. Klasifikasi dengan Algoritma *Naïve Bayes*.

Sekarang, kita akan menggunakan probabilitas yang telah dihitung untuk mengklasifikasikan data uji.

Tabel 3. 13 Klasifikasi Data Uji

Nama Karyawan	S1	S2	S3	S4	S5	Lembur	Kriteria
Andrianto Syahputra	1	0	0	0	0	1,00	Disiplin
Surati	0	0	1	0	0	0,00	Disiplin
Sunardi	1	0	0	0	0	0,00	Tidak Disiplin
Indra Bakti	1	1	0	0	0	1,00	Disiplin

1. Prediksi untuk Andrianto Syahputra ($S_1=1, S_2=0, S_3=0, S_4=0, S_5=0,$
Lembur=1):

- $P(\text{Disiplin} | \text{Data})$

$$\begin{aligned} P(\text{Disiplin} | \text{Data}) &= P(\text{Disiplin}) \times P(S_1=1 | \text{Disiplin}) \times P(S_2=0 | \\ &\text{Disiplin}) \times P(S_3=0 | \text{Disiplin}) \times P(S_4=0 | \text{Disiplin}) \times P(S_5=0 | \\ &\text{Disiplin}) \times P(\text{Lembur}=1 | \text{Disiplin}) \\ &= 0,44 \times 0 \times 1 \times 0,86 \times 1 \times 1 \times 0,29 \\ &= 0 \end{aligned}$$

- $P(\text{Tidak Disiplin} | \text{Data})$

$$\begin{aligned} P(\text{Tidak Disiplin} | \text{Data}) &= P(\text{Tidak Disiplin}) \times P(S_1=0 | \text{Tidak Disiplin}) \\ &P(S_2=0 | \text{Tidak Disiplin}) \times P(S_3=0 | \text{Tidak Disiplin}) \times P(S_4=0 | \text{Tidak} \\ &\text{Disiplin}) \times P(S_5=0 | \text{Tidak Disiplin}) \times P(\text{Lembur}=1 | \text{Tidak Disiplin}) \\ &= 0,56 \times 0,44 \times 0,67 \times 1 \times 0,78 \times 0,56 \times 0 = 0 \end{aligned}$$

Andrianto Syahputra :

$$P(\text{Disiplin} | \text{Data}) = 0 \implies P(\text{Tidak Disiplin} | \text{Data}) = 0$$

Klasifikasi: Disiplin

2. Prediksi untuk Surati ($S_1=0, S_2=0, S_3=1, S_4=0, S_5=0, \text{Lembur}=0$)

- $P(\text{Disiplin} | \text{Data})$:

$$\begin{aligned} P(\text{Disiplin} | \text{Data}) &= 0,44 \times 1 \times 0,14 \times 1 \times 1 \times 1 \times 0,43 \\ &= 0,026488 \end{aligned}$$

- $P(\text{Tidak Disiplin} | \text{Data})$:

$$\begin{aligned} P(\text{Tidak Disiplin} | \text{Data}) &= 0,56 \times 0,67 \times 1 \times 0 \times 0,78 \times 0,56 \times 1 \\ &= 0 \end{aligned}$$

Surati:

$$P(\text{Disiplin} \mid \text{Data}) = 0,026488 > P(\text{Tidak Disiplin} \mid \text{Data}) = 0$$

Klasifikasi: Disiplin

3. Prediksi untuk Sunardi (S1=1, S2=0, S3=0, S4=0, S5=0, Lembur=0)

- P(Disciplin | Data):

$$P(\text{Disiplin} \mid \text{Data}) = 0,44 \times 0 \times 1 \times 0,86 \times 1 \times 1 \times 0,43$$

$$= 0$$

- P(Tidak Disiplin | Data):

$$P(\text{Tidak Disiplin} \mid \text{Data}) = 0,56 \times 0,44 \times 0,67 \times 1 \times 0,78 \times 0,56 \times 1$$

$$= 0,072110$$

Sunardi:

$$P(\text{Disiplin} \mid \text{Data}) = 0 < P(\text{Tidak Disiplin} \mid \text{Data}) = 0,072110$$

Klasifikasi: **Tidak** Disiplin

4. Prediksi untuk Indra Bakti (S1=1, S2=1, S3=0, S4=0, S5=0, Lembur=1)

- P(Disciplin | Data):

$$P(\text{Disiplin} \mid \text{Data}) = 0,44 \times 0 \times 0 \times 0,86 \times 1 \times 1 \times 0,29$$

$$= 0$$

- P(Tidak Disiplin | Data):

$$P(\text{Tidak Disiplin} \mid \text{Data}) = 0,56 \times 0,44 \times 0,33 \times 1 \times 0,78 \times 0,56 \times 0$$

$$= 0$$

Indra Bakti:

$$P(\text{Disiplin} \mid \text{Data}) = 0 == P(\text{Tidak Disiplin} \mid \text{Data}) = 0$$

Klasifikasi: Disiplin

Hasil prediksi *Naïve Bayes* untuk daya uji diperlihatkan pada tabel berikut:

Tabel 3. 14 Prediksi Pakai *Naïve Bayes*

Nama Karyawan	S1	S2	S3	S4	S5	Lembur	Kriteria
Andrianto Syahputra	1	0	0	0	0	1,00	Disiplin
Surati	0	0	1	0	0	0,00	Disiplin
Sunardi	1	0	0	0	0	0,00	Tidak Disiplin
Indra Bakti	1	1	0	0	0	1,00	Disiplin

2.10 Evaluasi Kinerja

Setelah implementasi, model diuji dengan metrik evaluasi berikut: Akurasi (Accuracy), Presisi (Precision), Recall, dan F1-Score. Sebelum diuji dengan matrik evaluasi kinerja. Maka kita harus menyusun Confusion Matrix. Kita membandingkan hasil prediksi dengan data aktual pada Data Uji. Berikut adalah tabel perbandingan:

Tabel 3. 15 Perbandingan Hasil Kriteriaia Aktual vs Prediksi

Nama Karyawan	S1	S2	S3	S4	S5	Lembur	Kriteria (Aktual)	Kriteria (Prediksi)
Andrianto Syahputra	1	0	0	0	0	1,00	Disiplin	Disiplin
Surati	0	0	1	0	0	0,00	Tidak Disiplin	Disiplin
Sunardi	1	0	0	0	0	0,00	Tidak Disiplin	Tidak Disiplin
Indra Bakti	1	1	0	0	0	1,00	Disiplin	Disiplin

Berikut adalah Confusion Matrix yang telah dikonversi ke dalam format TP (True Positive), TN (True Negative), FP (False Positive), FN (False Negative) untuk masing-masing kelas Disiplin dan Tidak Disiplin.

Tabel 3. 16 Confusion Matrix

	Prediksi Disiplin	Prediksi Tidak Disiplin
Aktual Disiplin	2 (TP)	0 (FN)
Aktual Tidak Disiplin	1 (FP)	1 (TN)

Penjelasan:

- TP (True Positive) dengan Nilai 2: Ini menunjukkan bahwa terdapat 2 kasus di mana model dengan benar mengklasifikasikan karyawan sebagai "Disiplin". Artinya, model berhasil mengidentifikasi karyawan yang memang " Disiplin ".
- FN (False Negative) dengan Nilai 0: Ini menunjukkan bahwa tidak ada kasus di mana model salah mengklasifikasikan karyawan yang sebenarnya " Disiplin " sebagai "Tidak Disiplin ".
- FP (False Positive) dengan Nilai 1: Ini menunjukkan bahwa ada 1 kasus di mana model salah mengklasifikasikan karyawan yang sebenarnya "Tidak Disiplin" sebagai " Disiplin ".
- TN (True Negative) dengan Nilai 1: Ini menunjukkan bahwa ada 1 kasus di mana model dengan benar mengklasifikasikan karyawan sebagai "Tidak Disiplin ".

A. Menghitung Akurasi

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

$$Akurasi = \frac{2 + 1}{2 + 1 + 1 + 0} = \frac{3}{4} = 0,75 \text{ atau } 75\%$$

B. Menghitung Presisi

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

$$Precision = \frac{2}{2 + 1} = 0,67 \text{ atau } 67\%$$

C. Menghitung Recall

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

$$Recall = \frac{2}{2 + 0} = 1 \text{ atau } 100\%$$

D. Menghitung F1-Score

$$F1Score = 2 \times \frac{precision \times recall}{precision + recall}$$

$$F1Score = 2 \times \frac{0,67 \times 1}{0,67 + 1}$$

$$F1Score = \frac{1,34}{1,67} = 0,80 \text{ atau } 80\%$$

Tabel 3. 17 Matrik Evaluasi Kinerja

Matrik Evaluasi	Hasil
Akurasi	75%
Precision	67%
Recall	100%
F1-Score	80%

Hasil dari matriks evaluasi kinerja yang ditampilkan pada Tabel 3.8 menunjukkan bahwa model klasifikasi memiliki performa kurang baik. Berikut adalah deskripsi dari setiap metrik yang ditampilkan:

- Akurasi 75% artinya, secara keseluruhan model mampu memprediksi dengan benar sebanyak 75% dari seluruh data uji..
- Presisi 67%, menunjukkan bahwa seberapa akurat model saat memprediksi karyawan " Disiplin ". Dalam hal ini, sekitar 66,7% prediksi " Disiplin " memang benar.
- Recall 100%, menggambarkan kemampuan model dalam mendeteksi semua kasus " Disiplin " dengan benar sebanyak 100%.
- F1-Score 80%, artinya memberikan keseimbangan antara presisi dan recall.

Hasil evaluasi menunjukkan bahwa algoritma *Naïve Bayes* memiliki performa yang cukup baik dalam memprediksi kedisiplinan karyawan. Nilai akurasi sebesar 75% menandakan bahwa model telah mampu melakukan klasifikasi dengan cukup akurat pada data uji. Nilai recall yang sempurna (100%) menunjukkan bahwa semua karyawan yang tergolong " Disiplin " berhasil dikenali oleh model. Namun, nilai presisi yang masih 66,7% mengindikasikan adanya beberapa prediksi yang keliru, yaitu memprediksi karyawan "Tidak Disiplin " sebagai " Disiplin ". Hal ini menunjukkan bahwa meskipun model sangat sensitif terhadap karyawan Disiplin, masih ada ruang untuk meningkatkan kemampuan diskriminasi terhadap kategori "Tidak Disiplin ". Perbaikan bisa dilakukan dengan menambah jumlah data latih, menyeimbangkan jumlah kelas, atau melakukan seleksi fitur yang lebih baik.