

BAB III METODE PENELITIAN

3.1 Objek Penelitian

Objek penelitian ini adalah buah apel, dengan fokus pada klasifikasi kualitasnya. Penelitian ini bertujuan untuk mengembangkan sistem otomatis yang dapat mengklasifikasikan kualitas apel menggunakan metode SVM, guna meningkatkan akurasi dan efisiensi dibandingkan metode manual. Proses penelitian ini dilakukan secara digital dengan data yang diperoleh dari *Kaggle*.

3.1.1 Waktu Penelitian

Penelitian dilakukan dengan memproses data yang telah dikumpulkan. Pengumpulan data berlangsung dari tanggal 3 November 2024 sampai April 2025.

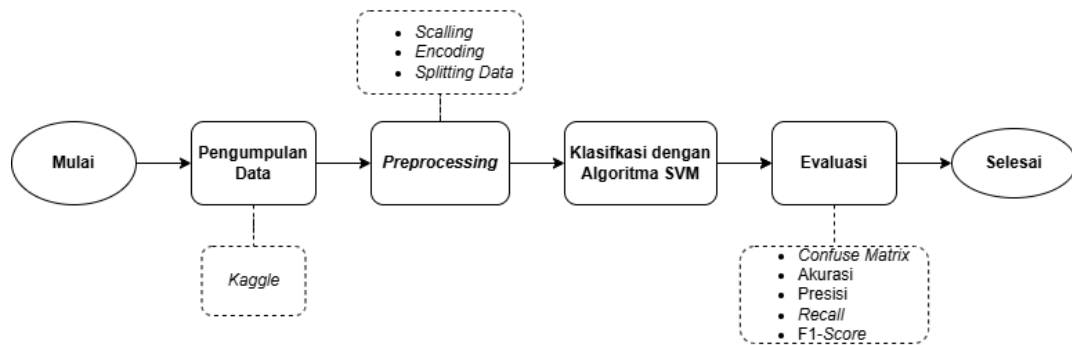
Tabel 3. 1 Waktu Penelitian

| Waktu Penelitian | Nov-24 | | | | Des-24 | | | | Jan-25 | | | |
|------------------------|--------|---|---|---|--------|---|---|---|--------|---|---|---|
| Tahapan Penelitian | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Pengumpulan Data | | | | | | | | | | | | |
| <i>Preprocessing</i> | | | | | | | | | | | | |
| <i>Splitting data</i> | | | | | | | | | | | | |
| Implementasi Algoritma | | | | | | | | | | | | |
| Evaluasi Model | | | | | | | | | | | | |

| Waktu Penelitian | Feb-25 | | | | Apr-25 | | | | Mei-25 | | | |
|------------------------|--------|---|---|---|--------|---|---|---|--------|---|---|---|
| Tahapan Penelitian | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Pengumpulan Data | | | | | | | | | | | | |
| <i>Preprocessing</i> | | | | | | | | | | | | |
| <i>Splitting data</i> | | | | | | | | | | | | |
| Implementasi Algoritma | | | | | | | | | | | | |
| Evaluasi Model | | | | | | | | | | | | |

3.2 Prosedur Penelitian

Untuk mencapai hasil yang diharapkan dalam penelitian, diperlukan prosedur penelitian yang jelas. Berikut prosedur yang dilakukan digambarkan dalam Gambar 3.1.



Gambar 3. 1 Prosedur Penelitian

Alur Penjelasan:

1. Pengumpulan *Dataset*: Data dikumpulkan dari repositori *daring Kaggle*. *Dataset* ini telah digunakan oleh penelitian sebelumnya oleh Astuti pada tahun 2024 menggunakan Algoritma KNN. *Dataset* terdiri dari 4000 *record* dan 9 atribut yang mencakup fitur *size* (ukuran), *weight* (berat), *sweetness* (tingkat kemanisan), *crunchiness* (kerenyahan), *juiciness* (kadar cairan), *ripeness* (kematangan), *acidity* (keasaman), *serta* label target *quality* (kualitas) (*bad / good*) untuk klasifikasi kualitas buah apel.

2. *Preprocessing*: Proses data untuk meningkatkan kualitasnya. Tahapan ini mencakup *scaling*, *encoding*.

a) *Scaling*

Scaling adalah proses penyesuaian nilai numerik agar berada pada skala yang konsisten. Penelitian ini menggunakan metode *StandardScaler*, yaitu mengubah nilai ke skala standar dengan mean 0 dan standar deviasi 1. Sebagai ilustrasi, data seperti *Size* dan *Weight* akan berubah nilainya setelah melalui proses *scaling*.

Tabel 3. 2 Contoh Sebelum Diterapkan Scaling

| Size | Weight |
|------|--------|
| 150 | 50 |
| 160 | 60 |
| 170 | 70 |

StandardScaler mengubah setiap nilai dengan rumus:

$$z = \frac{x - \mu}{\sigma} \quad (5)$$

Keterangan:

x = nilai asli

μ = rata-rata (*mean*) kolom

σ = standar deviasi (*standard deviation*) kolom

z = nilai setelah distandarkan

Tabel 3. 3 Contoh Setelah Diterapkan Scaling

| Size | Weight |
|-------|--------|
| -1.00 | -1.00 |
| 0.00 | 0.00 |
| 1.00 | 1.00 |

b) *Encoding*

Proses *encoding* adalah tahap mengubah data kategorikal (dalam bentuk teks) ke dalam bentuk numerik sehingga dapat dipahami dan diproses oleh algoritma *machine learning*. Dalam penelitian ini diterapkan metode Label *Encoding* yang merupakan salah satu teknik encoding paling sederhana. Teknik ini mengubah setiap nilai unik dalam kolom kategoris menjadi angka, di mana setiap kategori diwakili oleh bilangan bulat yang berbeda. Sebagai contoh, misalnya, pada data pelanggan terdapat kolom “Jenis_Kelamin” dengan nilai teks seperti “laki-laki” dan “Perempuan”. Melalui proses Label *Encoding*, nilai-nilai tersebut diubah menjadi representasi numerik.

Tabel 3. 4 Contoh Sebelum Diterapkan Label Encoding

| Id | Nama_Pelanggan | Jenis_Kelamin |
|-----------|-----------------------|----------------------|
| 1 | Mohan | Laki-laki |
| 2 | Raisa | Perempuan |
| 3 | Aqeela | Perempuan |
| 4 | Fattah | Laki-laki |
| 5 | Rony | Laki-laki |

Tabel 3. 5 Contoh Setelah Diterapkan Label Encoding

| Id | Nama_Pelanggan | Jenis_Kelamin |
|-----------|-----------------------|----------------------|
| 1 | Mohan | 0 |
| 2 | Raisa | 1 |
| 3 | Aqeela | 1 |
| 4 | Fattah | 0 |
| 5 | Rony | 1 |

Keterangan :

Laki – laki dikodekan sebagai 0

Perempuan dikodekan sebagai 1

Dengan demikian, kolom kategoris “Jenis_Kelamin” telah berhasil diubah menjadi bentuk numerik dan siap digunakan dalam pelatihan model.

3. *Splitting* data: Dataset dibagi menggunakan perbandingan 80:20, di mana 80% data digunakan untuk pelatihan dan 20% data untuk pengujian. Perbandingan ini dipilih karena ukuran dataset cukup besar (4000 data), memberikan keseimbangan antara data pelatihan dan pengujian, serta memastikan evaluasi model yang akurat.
4. Klasifikasi dengan Algoritma SVM: Model klasifikasi yang diterapkan dalam penelitian ini adalah algoritma *Machine Learning* yaitu *Support Vector Machine* (SVM) dengan menggunakan 3 kernel: *Linear*, *Radial Base Function* (RBF), dan *Polynomial*. 3 kernel tersebut digunakan pada pengujian oleh data latih sebagai pemilihan kernel terbaik yang akan digunakan pada pengujian oleh data uji. Algoritma SVM dipilih karena memiliki akurasi tinggi dan efektif untuk klasifikasi data dengan margin yang jelas. Pada tahap ini merupakan proses klasifikasi dengan algoritma SVM dengan menggunakan Bahasa pemrograman *python* dan dioperasikan di *google colab*.
5. Evaluasi: Proses mengukur dan menilai sejauh mana berfungsi dengan baik menggunakan metrik tertentu, seperti akurasi, *recall*, *precision* dan *F1-Score* adalah metrik. Untuk menghitung nilai metrik ini, matriks yang disebut *confusion matriks* (matriks kebingungan). Nilai dalam matriks ini digunakan untuk menghitung metrik evaluasi seperti akurasi, *recall*, dan presisi. Berikut

untuk menghitung evaluasinya :

a) Akurasi

Mengukur proporsi prediksi yang benar dari total prediksi.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Keterangan :

TP (*True Positive*) = Data positif yang diprediksi benar sebagai positif.

FN (*False Negative*) = Data positif namun diprediksi sebagai negatif.

FP (*False Positive*) = Data negatif namun diprediksi sebagai positif.

TN (*True Negative*) = Data negatif yang diprediksi benar sebagai negatif.

b) *Recall*

Mengukur kemampuan model mendeteksi data positif.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

Keterangan :

TP (*True Positives*) = Jumlah data positif yang diprediksi benar sebagai positif.

FN (*False Negatives*) = Jumlah data positif yang salah diprediksi sebagai negatif.

c) *Precision*

Mengukur akurasi prediksi positif dari seluruh prediksi positif.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (8)$$

Keterangan :

TP (*True Positives*) = Jumlah data positif yang diprediksi benar sebagai positif.

FP (*False Positives*) = Jumlah data negatif yang salah diprediksi sebagai positif.

d) *F1-Score*

Kombinasi harmonis antara precision dan recall untuk menangani ketidakseimbangan data.

$$F1 - Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \quad (9)$$

Keterangan :

Precision = Ketepatan model dalam memprediksi kelas positif.

Recall = Kemampuan model dalam mendeteksi semua data positif.

