



Jurnal Teknologi Informasi dan Komunikasi

Vol: 13 No 1 2022

E-ISSN: 2477-3255

Diterima Redaksi: 30-03-2022 | Revisi: 13-05-2022 | Diterbitkan: 27-05-2022

Quality Classification of Palm Oil Varieties Using Naive Bayes Classifier

Novianti Puspitasari¹, Fhanji Wilis Pratama², Rosmasari³, Heni Sulastri⁴

^{1,2,3}Program Studi Informatika Fakultas Teknik Universitas Mulawarman

⁴Jurusan Informatika Fakultas Teknik Universitas Siliwangi

^{1,2,3}Jl. Sambaliung No 9 Kampus Gn. Kelua, Samarinda, Kalimantan Timur

⁴Jl. Siliwangi No 24 Kahuripan Kota Tasikmalaya

e-mail: ¹novia.ftik.unmul@gmail.com, ²skrrr48@gmail.com, ³rosmasari.unmul@gmail.com, ⁴henisulastri@unsil.ac.id

Abstract

As one of the leading commodities of the Indonesian economy, the ever-increasing production of palm oil has created intense competition among palm oil (CPO) producers. This causes CPO producers to increase their palm oil production without compromising the quality of the palm oil produced. CPO producers are required to be able to objectively determine the quality of superior and precise oil palm varieties in order to produce high economic value palm oil. Therefore, a model is needed to determine the quality of oil palm from several existing varieties. The Naive Bayes Classifier method in this study was used to classify the quality of oil palm based on predetermined variables using a data set of 28 oil palm varieties. Method testing is done by using a confusion matrix and K-fold cross-validation scheme. This study shows a reasonably high accuracy value of 64.25% and a low error rate of 35.7%, indicating that the Naive Bayes Classifier can classify the quality of oil palm varieties quite well.

Keywords: *confusion matrix, naive bayes classifier, palm oil, quality, varieties*

Klasifikasi Kualitas Varietas Kelapa Sawit Menggunakan Metode Naive Bayes Classifier

Abstrak

Sebagai salah satu komoditas unggulan perekonomian Indonesia, produksi minyak kelapa sawit yang terus meningkat menimbulkan persaingan ketat diantara para produsen minyak kelapa sawit (CPO). Hal ini menyebabkan produsen CPO harus meningkatkan hasil produksi minyak kelapa sawit tanpa mengesampingkan kualitas dari minyak sawit yang dihasilkan. Produsen CPO dituntut untuk mampu menentukan kualitas varietas kelapa sawit yang unggul dan tepat secara objektif demi menghasilkan minyak sawit bernilai ekonomi tinggi. Oleh karena itu, dibutuhkan suatu model untuk menentukan kualitas kelapa sawit dari beberapa varietas yang ada. Metode Naive Bayes Classifier dalam penelitian ini digunakan untuk mengklasifikasikan kualitas kelapa sawit berdasarkan variabel yang telah ditetapkan dengan penggunaan data set sebanyak 28 varietas kelapa sawit. Pengujian metode dilakukan dengan menggunakan skema confusion matrix dan K-fold cross-validation. Hasil dari penelitian ini

menampilkan nilai akurasi yang cukup tinggi sebesar 64.25% dan error rate yang rendah sebesar 35.7%, menunjukkan bahwa Naive Bayes Classifier mampu mengklasifikasikan kualitas varietas kelapa sawit dengan cukup baik.

Kata kunci: confusion matrix, kelapa sawit, kualitas, naive bayes classifier, varietas

1. Pendahuluan

1.1. Latar Belakang

Kelapa sawit mempunyai peran strategis dalam perekonomian makro Indonesia, sehingga saat ini kelapa sawit mengalami peningkatan produksi cukup pesat dibandingkan dengan tanaman perkebunan lainnya. Sebagai salah satu penghasil devisa terbesar, kedaulatan energi dan juga sebagai penggerak perekonomian nasional serta pendorong sektor ekonomi kerakyatan dalam hal menyerap tenaga kerja, industri kelapa sawit sendiri telah berkembang dan menyebar ke daerah-daerah potensial di 22 provinsi dari total 33 provinsi yang ada di Indonesia. Kalimantan Timur merupakan salah satu provinsi sentra industri kelapa sawit di Indonesia dengan jumlah produksi minyak sawit (*Crude Palm Oil*) yang dihasilkan per tahun 2018 mencapai sekitar 13,398,363 ton. [1]–[3].

Produksi minyak kelapa sawit yang terus meningkat menimbulkan persaingan yang semakin ketat diantara para produsen minyak kelapa sawit (CPO) demi memenuhi permintaan pasar yang berasal dari luar maupun dalam negeri. Dalam memenuhi permintaan CPO, produsen CPO tidak hanya dituntut untuk meningkatkan kapasitas produksi, namun juga dituntut untuk meningkatkan kualitas CPO yang dihasilkan [4]. Sebagai faktor penting dalam kelangsungan bisnis industri minyak kelapa sawit, produsen CPO harus mampu menentukan kualitas varietas kelapa sawit yang unggul dan tepat secara objektif untuk menghasilkan minyak sawit yang bernilai ekonomi tinggi. Penentuan kualitas varietas kelapa sawit dapat dilakukan dengan melakukan klasifikasi. Hal ini dikarenakan klasifikasi dapat memberikan solusi yang objektif untuk menemukan kualitas varietas kelapa sawit yang saat ini tersedia dipasaran. Klasifikasi merupakan suatu cara menganalisis data dengan menggunakan *training data* (data latih) untuk membentuk sebuah model yang akan digunakan dalam memprediksi kelas label dari sebuah data. Klasifikasi melakukan pengelompokkan fitur ke dalam kelas yang sesuai [5]. Metode klasifikasi yang telah banyak digunakan diantaranya adalah *Neural networks*, *Analisa Statistik*, *Decision Tree*, *Naive Bayes classifier*, *Algoritma Genetika*, *Rough Sets*, *Support Vector Machines (SVM)*, *K-Nearest Neighbor*, dan *Memory Based Reasoning*.

Penelitian ini melakukan klasifikasi terhadap beberapa varietas kelapa sawit dengan cara memprediksi label kelas dari setiap varietas kelapa sawit yang akan diuji menggunakan salah satu metode klasifikasi yaitu *Naive Bayes Classifier*. Pemilihan metode *Naive Bayes* didasarkan pada nilai akurasi yang cukup tinggi pada berbagai data penelitian, diantaranya adalah klasifikasi terhadap data nasabah asuransi yang menghasilkan kesimpulan bahwa metode ini memiliki akurasi yang cukup baik berdasarkan percobaan yang telah dilakukan sebanyak dua kali yaitu sebesar 80% di percobaan pertama dan di percobaan kedua sebesar 74,67% [6]. Penelitian selanjutnya menggunakan metode *Naive Bayes* untuk mengklasifikasikan penyakit diabetes, status gizi, kinerja satpam, dan penerima kartu Indonesia sehat, dengan persentase akurasi diatas 85% [7]–[10]. Selain itu algoritma *Naive Bayes Classifier* mampu memiliki tingkat error minimal bila dibandingkan dengan algoritma klasifikasi lainnya [10].

Dari beberapa penelitian tersebut dan penelitian lainnya [11]–[15] menunjukkan bahwa metode *Naive Bayes* dapat digunakan dalam menyelesaikan permasalahan terkait klasifikasi, karena metode ini memiliki kemampuan yang cepat dalam membangun model, melakukan prediksi dan menyediakan metode baru dalam memahami maupun mempelajari suatu data. Selain itu, metode ini menggunakan perhitungan statistik dan probabilitas yang mampu memprediksi peluang dimasa depan berdasarkan pengalaman di masa sebelumnya dan memiliki tingkat keakuratan yang baik [16]. Hal inilah yang menjadikan *Naive Bayes* sebagai algoritma prediksi yang paling cepat untuk dipelajari [9].

Berdasarkan paparan tersebut, penelitian yang dilakukan yaitu melakukan prediksi (klasifikasi) kualitas tanaman kelapa sawit dengan menerapkan metode *Naïve Bayes Classifier* dengan tujuan untuk membantu produsen kelapa sawit dalam menentukan varietas kelapa sawit yang berkualitas agar dapat meningkatkan kualitas produksi minyak sawit yang dihasilkan.

1.2. State of The Art

1.2.1. Kelapa Sawit

Kelapa sawit sebagai tanaman andalan penghasil minyak nabati yang penting dalam dunia perdagangan Internasional, banyak digunakan oleh sebagian besar masyarakat untuk konsumsi non-makanan dan makanan. Hal ini dikarenakan kelapa sawit menghasilkan minyak dengan kadar kolesterol rendah dibandingkan dengan minyak nabati yang dihasilkan oleh tanaman lain seperti kacang tanah dan kacang kedelai. Selain itu, kelapa sawit juga dikenal sebagai tanaman yang paling tahan terhadap serangan hama dan penyakit dibandingkan dengan tanaman penghasil minyak nabati lainnya [17].

Kelapa sawit yang ada di Indonesia terdiri dari berbagai macam varietas, dari beberapa varietas tersebut, tentu terdapat varietas yang diunggulkan. Persilangan antara beberapa jenis tanaman kelapa sawit yang memiliki daya gabung baik dengan prinsip reproduksi dari hibrida dilakukan sebagai upaya untuk menghasilkan varietas unggul kelapa sawit. Selain itu, varietas unggul lebih mudah untuk dikembangkan dan memiliki hasil panen yang berlimpah. Pemilihan varietas yang unggul sangat berpengaruh dalam menghasilkan minyak kelapa sawit yang berkualitas.

Kualitas suatu produk, khususnya minyak yang dihasilkan dari kelapa sawit merupakan faktor yang sangat penting bagi keberlanjutan bisnis (industri). Minyak goreng kualitas super umumnya diantaranya dipengaruhi oleh warna serta *cloud point* dan kandungan *Free Fatty Acid* (FFA) yang biasanya terdapat pada varietas unggul kelapa sawit [18].

1.2.2. Klasifikasi

Klasifikasi merupakan suatu teknik yang didasarkan pada atribut dan tingkah laku dari sekelompok data yang telah didefinisikan. Klasifikasi memiliki empat komponen dasar yaitu training set, pengujian dataset, kelas, dan prediktor [19]. Konsep klasifikasi sendiri bertujuan untuk melakukan prediksi penentuan kelas pada setiap data. *Supervised induction* digunakan untuk penentuan kelas dengan memanfaatkan sekumpulan pengujian data set yang telah terklasifikasi. Didalam proses klasifikasi terdapat dua proses yaitu fase pembelajaran (*training*) dan fase kedua adalah pengujian (*testing*), proses pembelajaran dan pengujian akan berjalan sesuai dengan metode algoritma yang akan digunakan untuk klasifikasi [20].

1.2.3. Naive Bayes Classifier

Naive Bayes Classifier merupakan sebuah metode klasifikasi yang memprediksi peluang (probabilitas) suatu kelas. Cara yang digunakan adalah melakukan perhitungan terhadap sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dataset yang digunakan [6]. Ciri utama dari metode ini adalah jumlah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi kejadian. Metode ini mengasumsikan bahwa semua atribut independent atau tidak saling memiliki ketergantungan berdasarkan nilai yang diberikan pada variabel kelas. Metode ini memiliki tingkat akurasi yang tinggi dengan waktu komputasi yang cepat [21]. Tahapan penyelesaian dengan menerapkan metode *Naive Bayes Classifier* sebagai berikut:

1. Jika data berbentuk numerik, hitunglah nilai probabilitas dari setiap kategori yang sama, dengan jumlah data dari kategori yang sama kemudian dibagi dengan data pada kategori tersebut.
2. Jika terdapat data numerik, maka pada tahap pertama yaitu mencari nilai standar deviasi dan *mean* dari setiap parameter yang menggambarkan data angka pada data latih (*training*). Rumus untuk mencari nilai *mean* terlihat pada persamaan 1.

$$\mu = \frac{\sum_{i=1}^n X_i}{n} \quad (1)$$

Dimana:

μ : rata-rata hitung (mean)

x_i : nilai sampel ke-i

n : jumlah sampel

- Selanjutnya adalah menghitung nilai standar deviasi dengan cara menghitung nilai varian yang mempresentasikan seluruh populasi menggunakan persamaan 2.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{n}} \quad (2)$$

Dimana:

σ : standar deviasi

x_i : nilai x ke -i

μ : rata-rata hitung

n : jumlah sampel

- Tahap berikutnya adalah menghitung nilai probabilitas yang berada pada setiap fitur kelas dengan cara menjumlahkan jumlah data yang sesuai berdasarkan kategori yang sama kemudian dibagi dengan jumlah data pada kategori tersebut.
- Setelah mendapatkan nilai *mean* dan standar deviasi selanjutnya adalah menghitung distribusi *Gaussian* dengan menggunakan persamaan 3.

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \exp \frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \quad (3)$$

- Setelah semua tahap selesai maka tahap selanjutnya adalah membandingkan nilai probabilitas akhir pada setiap label kelas dengan cara memasukkan semua data nilai distribusi *Gaussian* yang ada ke dalam satu kelas yang sama.
- Proses selanjutnya yaitu menganalisa hasil yang didapat pada proses sebelumnya dengan cara mencari nilai probabilitas akhir yang paling besar. Nilai probabilitas akhir didapatkan berdasarkan perhitungan nilai probabilitas akhir kelas yang dimasukkan ke dalam persamaan *Naive Bayes Classifier*. Perhitungan untuk mencari nilai probabilitas akhir dapat dilihat pada persamaan (4).

$$P(\text{Kelas}|X) = P(\text{Kelas}) * P(X) \quad (4)$$

- Proses terakhir yaitu melakukan normalisasi terhadap nilai probabilitas yang memiliki satu kategori yang sama dengan jumlah nilai pada semua kategori dengan cara membagi nilai kategori tersebut seperti pada persamaan (5).

$$P(\text{Kelas}) = P(\text{Kelas}|X) / (P(X|\text{Kelas}) + P(X|\text{Kelas})) \quad (5)$$

1.2.4. Confusion Matrix

Metode yang dapat digunakan untuk mengukur kinerja model klasifikasi salah satunya adalah *Confusion matrix*. *Confusion matrix* bekerja dengan cara membandingkan hasil klasifikasi dengan penerapan algoritma yang dilakukan dengan hasil klasifikasi yang seharusnya sehingga menghasilkan informasi tingkat akurasi penerapan algoritma tersebut [22]. Tidak hanya nilai *accuracy*, *Confusion matrix* juga dapat menghasilkan persentase jumlah *record* data bernilai valid (benar) dalam prediksi oleh algoritma dan jumlah persentase *record* data prediksi yang bernilai salah oleh algoritma. Tabel *Confusion Matrix* dapat dilihat sesuai dengan Tabel 1 [23].

Berdasarkan Tabel 1, *True Positive* (TP) menunjukkan nilai positif berdasarkan jumlah *record* data positif yang diklasifikasi. *False Positives* (FP) menunjukkan nilai positif berdasarkan jumlah data *record* negatif yang diklasifikasi. *False Negative* (FN) menunjukkan nilai negative berdasarkan jumlah *record* data positif yang diklasifikasi, sedangkan *True Negative* (TN) menunjukkan nilai negative berdasarkan jumlah *record* data negatif yang diklasifikasi.

Tabel 1. Confussion Matrix

Klasifikasi Benar	Klasifikasi	
	Prediksi Benar	Prediksi Salah
Aktual Positif	True Positive	False Negative
Aktual Negatif	False Positive	True Negative

1.2.5. K-Fold Cross Validation

Untuk memvalidasi kinerja model atau algoritma klasifikasi dapat menggunakan metode *K-Fold Cross Validation* yang merupakan metode statistik yang bersifat prediksi secara silang. Metode ini memisahkan dataset sebanyak K bagian secara acak, kemudian proses klasifikasi akan dilakukan sebanyak k kali percobaan dan hasil evaluasi pun memiliki jumlah nilai K yang sama, dimana masing-masing percobaan menggunakan data partisi ke-K sebagai data proses pembelajaran (*training*) dan sisanya sebagai data validasi (*testing*). Metode *Confusion matrix* digunakan berdasarkan penelitian [24], [25] yang menganggap bahwa metode ini dapat mengurangi waktu komputasi dengan tetap menjaga keakuratan estimasi.

2. Metode Penelitian

Pada bab ini akan dijabarkan tentang tahapan, metode, dan data penelitian yang digunakan.

2.1. Tahapan Penelitian

Tahapan metode penyelesaian penelitian ini adalah sebagai berikut:

1. Identifikasi Masalah dan Studi Literatur

Pada tahap ini, dilakukan pengamatan terhadap masalah yang akan diselesaikan terkait data-data kriteria yang dapat digunakan sebagai parameter untuk menentukan kualitas varietas dari kelapa sawit. Analisis dilakukan dengan tujuan untuk mencari solusi terhadap permasalahan tersebut dan untuk menentukan ruang lingkup permasalahan yang akan diteliti. Setelah ruang lingkup permasalahan ditentukan, selanjutnya yaitu mempelajari berbagai teori dasar dari literatur-literatur tentang penerapan metode *Naive Bayes Classifier* sebagai metode yang akan digunakan dalam penelitian, konsep dan teori terkait data mining dan literatur terkait kualitas kelapa sawit, melalui jurnal-jurnal yang relevan.

2. Pengumpulan Data

Tahap pengumpulan data merupakan tahap inti dari suatu penelitian. Prosedur yang digunakan dalam tahap pengumpulan ini diperoleh langsung dari Pusat Penelitian Kelapa Sawit (PPKS) kota Samarinda dan juga melalui website Kementerian Pertanian, serta melakukan wawancara dengan pihak PPKS untuk memperoleh data variabel dan data varietas sawit yang akan digunakan dalam penelitian.

3. Pembersihan Data (*Cleaning Data*)

Data yang telah dikumpulkan selanjutnya dilakukan proses *cleaning* data (pembersihan data) dengan tujuan untuk menghindari duplikasi nilai dan variabel yang tidak sesuai. Pada penelitian ini variabel yang digunakan terdapat pada Tabel 2.

Tabel 2. Variabel Penelitian

No	Nama Variabel	Penjelasan
1.	Rerata Jumlah Tandan	Rata-rata jumlah tandan yang dihasilkan varietas setiap tahun & per pohon
2.	Rerata Berat Tandan	Rata-rata berat tandan yang dihasilkan varietas dalam satuan kilogram (kg)
3.	Rerata TBS	Rata-rata tandan buah segar yang dihasilkan varietas setiap tahun dalam satuan tonase
4.	Rendemen	Kandungan minyak yang dihasilkan varietas kelapa sawit dalam satuan %

Variabel yang terpilih didapatkan dari hasil wawancara kepada ahli dalam bidang kelapa sawit yang menjabat sebagai penanggung jawab di PPKS Outlet Samarinda. Keempat variabel tersebut merupakan variabel yang mempengaruhi kualitas kelapa sawit di Indonesia.

4. Transformasi Data

Tahap transformasi merupakan proses pengubahan data sesuai dengan aturan yang telah ditetapkan sebagai berikut:

- a. Varietas yang memiliki rerata jumlah tandan per pohon pada rentang 19-23 keatas, rerata berat tandan per kilogram pada rentang 12-14, rerata TBS per ton pada rentang 27-29, dan rendemen dalam satuan persen pada rentang 26-29 diberikan label kelas “sangat baik”.
- b. Varietas yang memiliki rerata jumlah tandan per pohon pada rentang 14-18, rerata berat tandan per kilogram pada rentang 10-12, rerata TBS per ton pada rentang 25-26, dan rendemen dalam satuan persen pada rentang 24-26 diberikan label kelas “baik”.
- c. Varietas yang memiliki rerata jumlah tandan per pohon pada rentang 11-13, rerata berat tandan per kilogram pada rentang 8-10, rerata TBS per ton pada rentang 21-24, dan rendemen dalam satuan persen pada rentang 21-23 diberikan label kelas “cukup baik”.

5. Implementasi

Tahap implementasi merupakan tahap pengolahan data menggunakan sebuah alat bantu (*tools*). Alat bantu yang akan digunakan dalam tahap ini adalah Matlab. Di tahap ini diterapkan pula algoritma prediksi data mining yang telah terpilih yaitu *Naive Bayes Classifier* pada data yang telah dikumpulkan.

6. Pengujian

Setelah melakukan tahap implementasi, dilakukan pengujian untuk memeriksa tingkat keakuratan dan memeriksa pola yang dihasilkan untuk melihat kesesuaian ataupun ketidaksesuaian pola dengan hipotesa yang ada sebelumnya. Pengujian pada penelitian ini menerapkan metode *Confusion Matrix* dan *K-Fold Cross Validation*.

2.2. Data Penelitian

Penggunaan data set dalam penelitian ini terbagi menjadi dua bagian yaitu data latih (*training*) dan data uji (*testing*), berjumlah 28 data profil varietas kelapa sawit. Data latih pada penelitian ini berupa data profil varietas kelapa sawit unggul di Indonesia yang sudah diberikan label kelas. Data ini memiliki empat variabel yang terlihat pada Tabel 3.

Tabel 3. Data Training

No	Varietas	Rerata Jumlah Tandan	Rerata Berat Tandan	Rerata TBS	Rendemen	Klasifikasi
1	D x P - TS 1	16	11,6	25,2	24,7	Baik
2	D x P - TS 2	16	11,8	25,2	24,6	Baik
3	D x P - TS 3	13,0	10	23,6	23	Cukup baik
4	DP SocfindoL	18,6	13	31,5	27,4	Sangat baik
5	DP Socfindo Y	9,9	22,3	29,5	26,8	Sangat baik
6	D x P SJ-1	16,3	12,4	25,4	25	Baik
7	D x P SJ-2	14,9	12,8	25,9	25,2	Baik
8	D x P SJ-3	16,5	11,8	26,4	24,5	Baik
9	D x P SJ-4	14,4	12,3	24	25	Baik
10	D x P SJ-5	15,2	11,9	24,5	23,8	Cukup baik
11	D x P SJ-6	14,9	12,2	24,6	26,2	Baik

<https://doi.org/10.31849/digitalzone.v13i1.9773>

12	AA-DP TOPAZ 1	28	7,6	28,7	25	Sangat baik
13	AA-DP TOPAZ 2	27	7,8	28	26,3	Sangat baik
14	AA-DP TOPAZ 3	28	7,4	28,1	24,8	Baik
15	AA-DP TOPAZ 4	24	7,9	25,1	26	Sangat baik
16	D x P Bah Lias 1	22	13	26,7	26,3	Sangat baik
17	D x P Bah Lias 1 (SK)	22	12,8	24,7	27,6	Sangat baik
18	D x P Bah Lias 3	22	11,1	25,6	25,8	Baik
19	D x P Bah Lias 4	12,7	10,6	24,1	23,6	Cukup baik
20	D x P Bah Lias 7	23	10,5	25,5	28,2	Sangat baik

Penentuan label kelas pada data latih ditentukan oleh persentase kandungan minyak yang dihasilkan dari setiap varietas, tandan buah segar yang dihasilkan varietas per tahunnya dan dari hasil wawancara kepada ahli kelapa sawit dan menjabat sebagai penanggung jawab di Pusat Penelitian Kelapa Sawit (PPKS).

Data uji (*testing*) pada penelitian ini berupa sampel data uji varietas kelapa sawit unggul yang diproduksi dan disebarakan oleh PPKS Samarinda dari tahun 2016 - 2018 yang belum memiliki label kelas. Proses *mining* dilakukan dan menghasilkan kelas data berdasarkan probabilitas akhir yang diperoleh, data *testing* penelitian ini ditampilkan oleh Tabel 4.

Tabel 4. Data Testing

No	Varietas	Rerata Jumlah Tandan	Rerata Berat Tandan	Rerata TBS	Rendemen	Klasifikasi
1	Dy x P SP-1 (Dumpy)	8	25	28	24,6	?
2	D x P Avros	12	16	27	24,8	?
3	D x P Simalungun	12,9	19,2	28,4	26,5	?
4	D x P PPKS 540	12	16	25	23,9	?
5	D x P Yangambi	13	16	28	24,8	?
6	D x P PPKS 718	9,3	22,8	26	23,9	?
7	D x P 239	15,3	17,2	32	25,8	?
8	D x P Langkat	12,5	19	27,5	26,3	?

Data uji pada penelitian ini terdiri dari delapan data varietas unggul kelapa sawit yang diproduksi dan disebarakan oleh PPKS. Kedelapan varian tersebut memiliki karakter masa produksi dan pelepas yang relatif pendek sehingga petani lebih banyak memilih varian-varian tersebut.

3. Hasil dan Pembahasan

Hasil dan pembahasan memuat penjelasan hasil penelitian terkait klasifikasi kualitas varietas kelapa sawit menggunakan metode *Naive Bayes Classifier* dan proses pengujian terhadap hasil klasifikasi yang telah dilakukan.

3.1. Klasifikasi *Naive Bayes Classifier*

Penelitian ini menggunakan data numerik sehingga tahap pertama adalah menghitung nilai *mean* dan standar deviasi pada data *training*. Hal pertama yang harus dilakukan sebelum mencari nilai *mean* dan standar deviasi adalah mengelompokkan jumlah label kelas yang sesuai dengan Tabel 3. Berdasarkan Tabel 3 di atas terlihat bahwa dalam penelitian ini menggunakan tiga kelas yang dikelompokkan kedalam label sangat baik, baik dan cukup baik dengan bobot kelas Sangat Baik = 8, kelas Baik = 9 dan kelas Cukup Baik = 3

a. Nilai Mean

Perhitungan nilai mean pada setiap label kelas mengikuti persamaan (1).

Nilai *mean* pada label kelas Sangat Baik untuk setiap variabel sebagai berikut:

$$\mu \text{ Rerata jumlah tandan (sangat baik)} = (18,6 + 9,9 + 28 + 27 + 24 + 22 + 22 + 23)/8$$

$$\mu \text{ Rerata jumlah tandan (sangat baik)} = 21,8125$$

$$\mu \text{ Rerata berat tandan (sangat baik)} = (13 + 22,3 + 7,6 + 7,8 + 7,9 + 13 + 12,8 + 10,5)/8$$

$$\mu \text{ Rerata berat tandan (sangat baik)} = 11,8625$$

$$\mu \text{ Rerata TBS (sangat baik)} = (31,5 + 29,5 + 28,7 + 28 + 25,1 + 26,7 + 24,7 + 25,5)/8$$

$$\mu \text{ Rerata TBS (sangat baik)} = 27,4625$$

$$\mu \text{ Rendemen (sangat baik)} = (27,4 + 26,8 + 25 + 26,3 + 26 + 26,3 + 27,6 + 28,2)/8$$

$$\mu \text{ Rendemen (sangat baik)} = 26,7$$

Perhitungan nilai μ dengan label kelas Baik dan Cukup Baik pada variabel yang lainnya menggunakan langkah yang sama seperti perhitungan nilai *mean* pada label Sangat Baik. Nilai *mean* keseluruhan ditunjukkan Tabel 5 berikut.

Tabel 5. Nilai Mean

Klasifikasi	Rerata Jumlah Tandan	Rerata Berat tandan	Rerata TBS	Rendemen %
Sangat Baik	21,81	11,86	27,46	26,70
Baik	17,67	11,49	25,60	25,09
Cukup Baik	13,63	10,83	24,07	23,47

b. Standar Deviasi

Standar deviasi dilakukan setelah mendapatkan nilai μ pada masing-masing variabel.

Perhitungan nilai standar deviasi pada setiap label kelas mengikuti persamaan (2).

Nilai standar deviasi pada label kelas Sangat Baik sebagai berikut:

$$\sigma \text{ Rerata jumlah tandan (sangat baik)} = \sqrt{\frac{(((18,6 - 21,81)^2) + ((9,9 - 21,81)^2) + ((28 - 21,81)^2) + ((27 - 21,81)^2) + ((24 - 21,81)^2) + ((22 - 21,81)^2) + ((22 - 21,81)^2) + ((23 - 21,81)^2))}{(8 - 1)}}$$

$$\sigma \text{ Rerata jumlah tandan (sangat baik)} = \sqrt{223,6888/7}$$

$$\sigma \text{ Rerata jumlah tandan (sangat baik)} = 5,652923389$$

$$\sigma \text{ Rerata berat tandan (sangat baik)} = \sqrt{\frac{(((13 - 11,86)^2) + ((22,3 - 11,86)^2) + ((7,6 - 11,86)^2) + ((7,8 - 11,86)^2) + ((7,9 - 11,86)^2) + ((13 - 11,86)^2) + ((12,8 - 11,86)^2) + ((10,5 - 11,86)^2))}{(8 - 1)}}$$

$$\sigma \text{ Rerata berat tandan (sangat baik)} = \sqrt{164,6388/7}$$

$$\sigma \text{ Rerata berat tandan (sangat baik)} = 4,849724587$$

$$\sigma \text{ Rerata TBS (sangat baik)} = \sqrt{\frac{(((31,5 - 27,46)^2) + ((29,5 - 27,46)^2) + ((28,7 - 27,46)^2) + ((28 - 27,46)^2) + ((25,1 - 27,46)^2) + ((26,7 - 27,46)^2) + ((24,7 - 27,46)^2) + ((25,5 - 27,46)^2))}{(8 - 1)}}$$

$$\sigma \text{ Rerata TBS (sangat baik)} = \sqrt{39,9188/7}$$

$$\sigma \text{ Rerata TBS (sangat baik)} = 2,388029671$$

$$\sigma \text{ Rendemen (sangat baik)} = \sqrt{\frac{(((27,4 - 26,70)^2) + ((26,8 - 26,70)^2) + ((25 - 26,70)^2) + ((26,3 - 26,70)^2) + ((26 - 26,70)^2) + ((26,3 - 26,70)^2) + ((27,6 - 26,70)^2) + ((28,2 - 26,70)^2))}{(8 - 1)}}$$

$$\sigma \text{ Rendemen (sangat baik)} = \sqrt{7,2672/7}$$

$$\sigma \text{ Rendemen (sangat baik)} = 1,018906977$$

Perhitungan nilai σ dengan label kelas Baik dan Cukup Baik pada variabel yang lainnya menggunakan langkah yang sama seperti perhitungan nilai standar deviasi pada label Sangat Baik. Nilai standar deviasi keseluruhan ditampilkan oleh Tabel 6.

Tabel 6. Nilai Standar Deviasi

Klasifikasi	Rerata Jumlah Tandan	Rerata Berat tandan	Rerata TBS	Rendemen %
Sangat Baik	5,65	4,85	2,39	1,02
Baik	4,48	1,61	1,17	0,57
Cukup Baik	1,37	0,97	0,45	0,42

c. Probabilitas Setiap Kelas

Setelah mendapatkan nilai *mean* dan standar deviasi, maka langkah berikutnya adalah mencari nilai probabilitas pada data latih untuk setiap kelas. Perhitungan nilai probabilitas setiap kelas dilakukan dengan cara menghitung jumlah data yang sesuai dari setiap kategori. Di dalam penelitian ini, jumlah data yang digunakan adalah sebanyak 20 data. Data kelas “Sangat Baik” sebanyak 8 data, kelas “Baik” sebanyak 9 data, dan kelas “Cukup Baik” sebanyak 3 data. Perhitungan nilai probabilitas kelas Sangat Baik diuraikan sebagai berikut:

$$P(\text{Sangat baik}) = 8/20$$

$$P(\text{Sangat baik}) = 0,4$$

Hasil keseluruhan nilai probabilitas ditunjukkan pada Tabel 7 berikut.

Tabel 7. Nilai Probabilitas Setiap Kelas

Klasifikasi	Jumlah Data	Nilai Probabilitas
Sangat baik	8	0,4
Baik	9	0,45
Cukup baik	3	0,15
Jumlah	20	1

d. Nilai Distribusi *Gaussian*

Proses berikut yang harus dilakukan adalah menghitung nilai distribusi *Gaussian* berdasarkan pada data *testing* (Tabel 4). Data penelitian ini adalah data numerik, oleh sebab itu nilai probabilitas harus dihitung secara menyeluruh sesuai dengan data *testing* yang digunakan. Perhitungan *Gaussian* akan menggunakan data *Testing* $Dy \times P$ SP-1 (Dumpy) sebagai data sampel. Perhitungan menggunakan persamaan (3).

Nilai distribusi *Gaussian* pada label kelas Sangat Baik sebagai berikut:

$$P(\text{Rerata jumlah tandan} = 8 \mid \text{Sangat baik}) = \frac{1}{\sqrt{2\pi * 5,652923389}} * e^{-\frac{(8-21,8125)^2}{(2 * 31,95554285)}}$$

$$P(\text{Rerata jumlah tandan} = 8 \mid \text{Sangat baik}) = \frac{1}{14,16618540} * 2,718281^{-2,98516531}$$

$$P(\text{Rerata jumlah tandan} = 8 \mid \text{Sangat baik}) = 0,003567029$$

$$P(\text{Rerata berat tandan} = 25 \mid \text{Sangat baik}) = \frac{1}{\sqrt{2\pi * 4,849724587}} * e^{-\frac{(25-11,8625)^2}{(2 * 23,51982857)}}$$

$$P(\text{Rerata berat tandan} = 25 \mid \text{Sangat baik}) = \frac{1}{12,15337498} * 2,718281^{-3,669114886}$$

$$P(\text{Rerata berat tandan} = 25 \mid \text{Sangat baik}) = 0,082281670$$

$$P(\text{Rerata TBS} = 28 \mid \text{Sangat baik}) = \frac{1}{\sqrt{2\pi * 2,388029671}} * e^{-\frac{(28-27,4625)^2}{(2 * 5,702685714)}}$$

$$P(\text{Rerata TBS} = 28 \mid \text{Sangat baik}) = \frac{1}{5,984385202} * 2,718281^{-0,025330718}$$

$$P(\text{Rerata TBS} = 28 \mid \text{Sangat baik}) = 0,162921902$$

$$P(\text{Rendemen} = 24.6 \mid \text{Sangat baik}) = \frac{1}{\sqrt{2\pi \cdot 1.018906977}} * e^{-\frac{(24.6-26.7)^2}{(2 * 1.038171428)}}$$

$$P(\text{Rendemen} = 24,6 \mid \text{Sangat baik}) = \frac{1}{2,553373565} * 2,718281^{-2,12392668}$$

$$P(\text{Rendemen} = 24,6 \mid \text{Sangat baik}) = 0,046824838$$

Perhitungan nilai *Gaussian* dengan label kelas Baik dan Cukup Baik pada variabel lainnya dan data *testing* lainnya menggunakan langkah yang sama seperti perhitungan nilai *Gaussian* pada label Sangat Baik.

e. Probabilitas Akhir Setiap Kelas

Perhitungan nilai probabilitas akhir setiap kelas, dilakukan dengan cara mengelompokkan keseluruhan data dari nilai distribusi *Gaussian* yang ada kedalam satu kelas yang sama. Perhitungan ini menggunakan data *testing* Dy x P SP-1 (Dumpy) sebagai data sampel.

$$P(X \mid \text{Sangat baik}) = 0,003567029 * 0,082281670 * 0,162921902 * 0,046824838$$

$$P(X \mid \text{Sangat baik}) = 0,00000223905$$

Nilai yang didapatkan pada perhitungan diatas adalah probabilitas akhir untuk kelas dengan hasil Sangat Baik. Pencarian nilai probabilitas kelas dengan hasil Baik dan Cukup Baik, langkah-langkahnya sama seperti mencari nilai probabilitas akhir kelas dengan hasil Sangat Baik.

f. Probabilitas Akhir

Tahap berikutnya adalah menghitung probabilitas akhir yang didapatkan setelah memperoleh nilai probabilitas akhir setiap kelas. Perhitungan untuk mencari nilai probabilitas akhir untuk data *testing* Dy x P SP-1 (Dumpy) terlihat seperti berikut:

$$P(\text{Sangat baik} \mid X) = 0,4 * 0,00000223905$$

$$P(\text{Sangat baik} \mid X) = 0,00000089562350$$

Setelah melakukan proses perhitungan untuk mendapatkan nilai probabilitas akhir, proses terakhir adalah melakukan normalisasi terhadap seluruh kategori seperti berikut:

$$P(\text{Sangat baik}) = P(\text{Sangat baik} \mid X) / (P(X \mid \text{Sangat baik}) + P(X \mid \text{Baik}) + P(X \mid \text{Cukup baik}))$$

$$P(\text{Sangat baik}) = 0,00000089562350 / (0,00000223905 + 0 + 0)$$

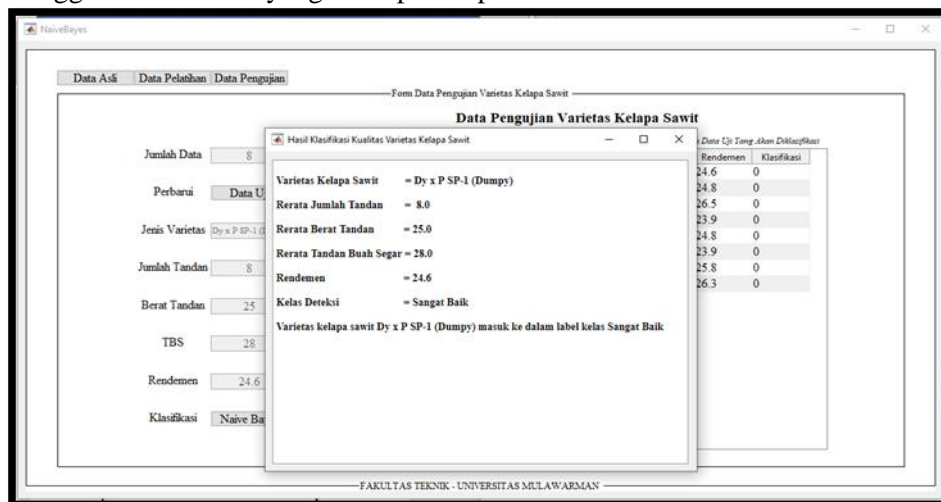
$$P(\text{Sangat baik}) = 0,40000156316$$

Nilai perhitungan probabilitas akhir untuk kelas Sangat Baik pada data sampel Dy x P SP-1 (Dumpy) adalah 0,40000156316. Proses untuk mendapatkan nilai probabilitas akhir dengan hasil Baik dan Cukup Baik dapat dilakukan dengan cara yang sama seperti menghitung nilai probabilitas akhir dengan hasil Sangat Baik. Keseluruhan hasil perhitungan probabilitas akhir ditunjukkan pada Tabel 8 berikut.

Tabel 8. Nilai Probabilitas Akhir

No	Varietas	Kelas Sangat Baik	Kelas Baik	Kelas Cukup baik	Klasifikasi
1	Dy x P SP-1 (Dumpy)	0,40000156316	0,00000000000	0,00000000000	Sangat Baik
2	D x P Avros	0,13305455562	0,30031360846	0,00000000000	Baik
3	D x P Simalungun	0,39999899097	0,00000113515	0,00000000000	Sangat Baik
4	D x P PPKS 540	0,05776146927	0,3848164785	0,00668616868	Baik
5	D x P Yangambi	0,26950363505	0,14680787124	0,00000000000	Sangat Baik
6	D x P PPKS 718	0,40000000000	0,00000000000	0,00000000000	Sangat Baik
7	D x P 239	0,39999983182	0,00000000000	0,00000000000	Sangat Baik
8	D x P Langkat	0,39998695947	0,00001467059	0,00000000000	Sangat Baik

Hasil dari perhitungan akhir juga sejalan dengan perhitungan yang telah dilakukan menggunakan Matlab yang ditampilkan pada Gambar 1.



Gambar 1. Hasil Klasifikasi

Berdasarkan hasil perhitungan yang telah dilakukan dan ditampilkan oleh Tabel 8, terdapat dua varietas yang memiliki kualitas Baik yaitu varietas D x P Avros dan D x P PPKS 540. Sedangkan enam varietas lainnya memiliki kualitas Sangat Baik.

3.2. Pengujian Metode

Skema *K-fold Cross Validation* digunakan untuk melakukan pengujian terhadap hasil dari metode klasifikasi yang digunakan. Dataset yang digunakan akan dibagi kedalam beberapa bagian secara acak. *fold* K-1, ketika bagian ke 1 menjadi data uji dan sisanya menjadi data latih, demikian seterusnya hingga sampai *fold* K-4. Pada evaluasi menggunakan *confusion matrix* melibatkan perhitungan *accuracy*, *precision*, *recall* dan *error rate*. Dataset dibagi menjadi *4-fold* dan masing-masing *fold* memiliki 7 data dengan persentase sebesar 75% untuk data training dan 25% untuk data uji. Nilai keseluruhan dari hasil pengujian pada semua *fold* ditunjukkan pada Tabel 9 berikut.

Tabel 9. Nilai Pengujian Metode

Dataset	Accuracy	Precision	Recall	Error rate
K-1	57,1%	50%	46,7%	42,8%
K-2	57,1%	75%	80%	42,8%
K-3	71,4%	33,3%	27,7%	28,6%
K-4	71,4%	33,3%	23,7%	28,6%
Rata-rata	64,25%	47,9%	44,5%	35,7%

Berdasarkan Tabel 9 diketahui bahwa pengujian tingkat *accuracy*, *precision* dan *recall* yang telah dilakukan menggunakan empat data set menghasilkan rata-rata nilai *accuracy* sebesar 64.25%, *precision* 47,9%, *recall* sebesar 44,5% dan *error rate* sebesar 35,7%. Dari hasil pengujian tingkat *accuracy* yang telah dilakukan terlihat bahwa persentase akurasi cenderung meningkat. Hal ini dikarenakan dengan adanya data pelatihan yang banyak tentu akan mendapatkan hasil yang akurat. Selain itu tingkat *error* yang dimiliki dibawah 50%. Hal ini menunjukkan bahwa metode ini memiliki *error* yang kecil.

4. Kesimpulan

Penerapan metode *Naive Bayes Classifier* yang telah dilakukan dalam penelitian ini menunjukkan bahwa metode tersebut dapat melakukan klasifikasi kualitas dari beberapa varietas kelapa sawit. Hasil klasifikasi menunjukkan bahwa terdapat dua varietas yang termasuk ke dalam kualitas “Baik” yaitu D x P Avros dan D x P PPKS 540. Sedangkan, enam varietas lainnya memiliki kualitas “Sangat Baik”. Klasifikasi didasarkan pada nilai probabilitas untuk

setiap kriteria pada masing-masing kelas yang berbeda. Nilai probabilitas dari kriteria tersebut dilakukan optimalisasi untuk mengidentifikasi kualitas kelapa sawit berdasarkan proses klasifikasi yang dilakukan oleh metode *Naive Bayes Classifier* tersebut. Lebih lanjut, dari hasil perhitungan *Naive Bayes Classifier* terhadap data varietas kelapa sawit menghasilkan persentase keakuratan sebesar 64,25% yang menunjukkan bahwa metode ini mampu mengklasifikasikan kualitas kelapa sawit dengan baik. Penelitian kedepannya diharapkan dapat menambah dataset dengan atribut yang lebih bervariasi dan penggunaan metode klasifikasi ataupun metode lainnya untuk mengetahui metode yang lebih tepat dan akurat dalam mengklasifikasikan kualitas kelapa sawit.

Daftar Pustaka

- [1] N. Ngadi and M. Noveria, "Keberlanjutan Perkebunan Kelapa Sawit Di Indonesia Dan Prospek Pengembangan Di Kawasan Perbatasan," *Masy. Indones.*, vol. 43, no. 1, 2018.
- [2] R. Advent, Z. Zulgani, and N. Nurhayani, "Analisis faktor-faktor yang mempengaruhi ekspor minyak kelapa sawit di Indonesia Tahun 2000-2019," *e-Journal Perdagang. Ind. dan Monet.*, vol. 9, no. 1, pp. 49–58, 2021.
- [3] J. H. V Purba and T. Sipayung, "Perkebunan kelapa sawit indonesia dalam perspektif pembangunan berkelanjutan," *Masy. Indones.*, vol. 43, no. 1, 2018.
- [4] G. Alfikri and N. L. P. Hariastuti, "Peningkatan Kualitas Minyak Kelapa Sawit Dengan Pendekatan Lean Six Sigma (Studi Kasus di PT. Sawit Mas Parenggean)," *J. Iptek*, vol. 23, no. 1, pp. 47–54, 2019.
- [5] A. P. Wibawa, "Metode-metode Klasifikasi," in *Prosiding SAKTI (Seminar Ilmu Komputer dan Teknologi Informasi)*, 2018, vol. 3, no. 1, pp. 134–138.
- [6] D. A. Novilla, R. Goejantoro, and F. D. T. Amijaya, "Klasifikasi Data Nasabah Asuransi Dengan Menggunakan Metode Naive Bayes," *EKSPONENSIAL*, vol. 10, no. 2, pp. 95–102, 2020.
- [7] A. P. Wibowo and S. Hartati, "Sistem Klasifikasi Kinerja Satpam Menggunakan Metode Naive Bayes Classifier," *INOVTEK Polbeng-Seri Inform.*, vol. 1, no. 2, pp. 192–201, 2016.
- [8] A. A. Rahman and Y. I. Kurniawan, "Aplikasi Klasifikasi Penerima Kartu Indonesia Sehat Menggunakan Algoritma Naive Bayes Classifier," *J. Teknol. dan Manaj. Inform.*, vol. 4, no. 1, 2018.
- [9] A. Ridwan, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020.
- [10] M. Y. Titimeidara and W. Hadikurniawati, "Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita," *J. Ilm. Inform.*, vol. 9, no. 01, pp. 54–59, 2021.
- [11] C. Fadlan, S. Ningsih, and A. P. Windarto, "Penerapan Metode Naive Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra," *JUTIM (Jurnal Tek. Inform. Musirawas)*, vol. 3, no. 1, pp. 1–8, 2018.
- [12] T. Imandasari, E. Irawan, A. P. Windarto, and A. Wanto, "Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air," in *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 2019, vol. 1, pp. 750–761.
- [13] T. A. Pratiwi, M. Irsyad, R. Kurniawan, S. Agustian, and B. S. Negara, "Klasifikasi Kebakaran Hutan Dan Lahan Menggunakan Algoritma Naive Bayes Di Kabupaten Pelalawan," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 6, no. 1, pp. 139–148, 2021.
- [14] V. R. S. Nastiti, S. Basuki, and H. Hilman, "Klasifikasi sinopsis novel menggunakan metode naive bayes classifier," *J. Repos.*, vol. 1, no. 2, pp. 125–130, 2019.
- [15] A. Natuzzuhriyyah, N. Nafisah, and R. Mayasari, "Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Secara Daring Menggunakan Algoritma Naive Bayes," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 6, no. 3, pp. 161–170, 2021.

-
- [16] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018.
- [17] D. Rahayu, R. C. Wihandika, and R. S. Perdana, "Implementasi Metode Backpropagation Untuk Klasifikasi Kenaikan Harga Minyak Kelapa Sawit," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN*, vol. 2548, p. 964X, 2018.
- [18] T. A. Mulyati, F. E. Pujiono, and P. A. Lukis, "Pengaruh lama pemanasan terhadap kualitas minyak goreng kemasan kelapa sawit," *J. Wiyata Penelit. Sains dan Kesehat.*, vol. 2, no. 2, pp. 162–168, 2017.
- [19] A. H. Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris," *J. Ilm. Ilmu Komput. Fak. Ilmu Komput. Univ. Al Asyariah Mandar*, vol. 7, no. 2, pp. 45–51, 2021.
- [20] I. Mubarog, A. Setyanto, and H. Sismoro, "Sistem Klasifikasi Pada Penyakit Breast Cancer Dengan Menggunakan Metode Naive Bayes," *Creat. Inf. Technol. J.*, vol. 6, no. 2, pp. 109–118, 2021.
- [21] R. Ghaniy and K. Sihotang, "Penerapan Metode Naive Bayes Classifier untuk Penentuan Topik Tugas Akhir," *Teknois J. Ilm. Teknol. Inf. dan Sains*, vol. 9, no. 1, pp. 63–72, 2019.
- [22] A. Ahmad and A. Latief, "Perbandingan Metode KNN Dan LBPH Pada Klasifikasi Daun Herbal," *J. RESTI (Rekayasa Sist. Dan Teknol. Informasi)*, vol. 5, no. 3, pp. 557–564, 2021.
- [23] M. F. Rahman, D. Alamsah, M. I. Darmawidjadja, and I. Nurma, "Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN)," *J. Inform.*, vol. 11, no. 1, p. 36, 2017.
- [24] S. T. Rizaldi, A. Al Khairi, and M. Mustakim, "Text Mining Classification Opini Publik Terhadap Provider di Indonesia," in *Seminar Nasional Teknologi Informasi Komunikasi dan Industri*, 2022, pp. 79–86.
- [25] M. D. Purbolaksono, M. I. Tantowi, A. I. Hidayat, and A. Adiwijaya, "Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes," *J. RESTI (Rekayasa Sist. Dan Teknol. Informasi)*, vol. 5, no. 2, pp. 393–399, 2021.
-