

Implementasi K-Nearest Neighbor Untuk Klasifikasi Jenis Kelamin Berdasarkan Analisis Citra Wajah

Maulana Fansyuri, Devi Yunita*

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: ¹dosen02359@unpam.ac.com, ²*dosen00846@unpam.ac.com

Email Penulis Korespondensi: dosen00846@unpam.ac.com

Abstrak-Wajah adalah bagian dari kepala, pada manusia meliputi area dari dahi hingga dagu, meliputi rambut, dahi, alis mata, mata, hidung, pipi, mulut, bibir, gigi, kulit dan dagu. Wajah digunakan untuk ekspresi wajah, penampilan dan identitas. Deteksi wajah adalah tahap untuk identifikasi pribadi, sistem pemantauan, hukum pidana, interaksi manusia-komputer. Era teknologi yang terus berkembang menuntut perkembangan dunia teknologi untuk menemukan pengetahuan teknologi baru yang lebih akurat dan cepat serta banyaknya permasalahan di bidang keamanan teknologi dan hukum pidana yang membutuhkan identifikasi klasifikasi wajah dalam menyelesaikan suatu masalah. Sistem pengenalan wajah membutuhkan suatu fitur dari suatu citra untuk dapat dikenali kemudian fitur tersebut akan dicocokkan dengan fitur citra lainnya. prosesnya membutuhkan metode ekstraksi ciri atau fitur dan klasifikasi K-Nearest. Tujuan dari penelitian ini adalah untuk mendapatkan tingkat akurasi terbaik dalam proses klasifikasi citra wajah dalam menentukan jenis kelamin seseorang. Metode dalam penelitian ini dilakukan dalam dua tahap, yaitu tahap pelatihan dan tahap pengujian. Pada tahap pelatihan, langkah-langkah yang dilakukan bertujuan untuk mendapatkan model berdasarkan subset citra yang disebut citra latih. Langkah awal penelitian adalah menyiapkan kumpulan data citra yang akan dianalisis. Dataset citra yang digunakan sebanyak 20 citra wajah kemudian diambil 10 citra. Berdasarkan hasil penelitian yang dilakukan terhadap citra wajah berdasarkan warna dan bentuk dengan menggunakan metode K-Nearest Neighbor, dapat disimpulkan bahwa metode ini termasuk dalam algoritma yang sangat baik untuk diterapkan pada klasifikasi citra wajah berdasarkan warna dan bentuk secara akurat. nilai 96%, sehingga penentuan jenis kelamin berdasarkan objek wajah menggunakan data yang diekstraksi dari warna dan bentuk serta menggunakan metode klasifikasi K-Nearest Neighbor sesuai dengan data citra sebenarnya.

Kata kunci: K-Nearest Neighbor; Wajah; Klasifikasi; Citra; Image Processing

Abstract -The face or face is part of the head, in humans it covers the area from forehead to chin, including hair, forehead, eyebrows, eyes, nose, cheeks, mouth, lips, teeth, skin and chin. The face is used for facial expressions, appearance and identity. Face detection is a stage for personal identification, monitoring systems, criminal law, human-computer interaction. The ever-evolving technological era demands the development of the world of technology to find new, more accurate and fast technological knowledge as well as many problems in the field of technological security and criminal law that require identification of facial classifications in solving a problem. The facial recognition system requires a feature from an image to be recognized then this feature will be matched with another image feature. the process requires a feature extraction method or features and K-Nearest classification. The purpose of this research is to get the best level of accuracy in the facial image classification process in determining a person's gender. The method in this study was carried out in two phases, namely the training phase and the testing phase. In the training phase, the steps taken aim to obtain a model based on a subset of images called training images. The initial step of research is to prepare image data sets to be analyzed. The image dataset used is 20 facial images and then takes 10 images. Based on the results of research conducted on facial images based on color and shape using the K-Nearest Neighbor method, it can be concluded that this method is included in an excellent algorithm for application to facial image classification based on color and shape with an accuracy value of 96%, so that the determination gender based on facial objects using data extracted from color and shape and using the K-Nearest Neighbor classification method according to the actual image data.

Keywords: K-Nearest Neighbor; Face; Classification; Image; Image Processing

1. PENDAHULUAN

Citra digital merupakan citra yang dapat diolah atau diproses oleh komputer, Pengolahan citra merupakan cara atau proses mengolah citra dengan berbagai teknik tertentu[1]. Teknik olah citra tersebut meliputi modifikasi kecermerlangan, negasi, peningkatan kontras, dan thresholding. Salah satu tujuan dari pengolahan citra adalah untuk pengenalan sebuah objek[2]. Pengenalan objek pada citra digital mendapat banyak perhatian dalam beberapa tahun terakhir ini. Hal ini terlihat dari penerapan identifikasi citra di beberapa bidang kerja seperti pengenalan gambar untuk perencanaan lahan pertanian, Biometrik dan pencitraan medis[3].

Wajah atau muka adalah bagian dari kepala, pada manusia meliputi wilayah dari dahi hingga dagu, termasuk rambut, dahi, alis, mata, hidung, pipi, mulut, bibir, gigi, kulit dan dagu[4]. Wajah digunakan untuk ekspresi wajah, penampilan serta identitas[5]. Manusia memiliki aneka bentuk wajah dan warna kulit yang berbeda, sehingga dapat digunakan dalam menentukan identitas seseorang. Pengenalan pola wajah dapat dipandang untuk mengenali emosi, ras, ataupun pemilik wajah tersebut berdasarkan fitur-fitur wajah yang dimiliki[6].

Deteksi wajah merupakan suatu tahapan untuk identifikasi personal, sistem pemantau, hukum kriminal, interaksi manusia dan komputer[1]. Era teknologi yang terus berkembang menuntut perkembangan dunia teknologi untuk menemukan ilmu teknologi yang baru, lebih akurat dan cepat serta banyak permasalahan dalam bidang kewanman teknologi dan hukum kriminal yang membutuhkan identifikasi klasifikasi wajah dalam menyelesaikan sebuah permasalahan[7].

Sistem pengenalan wajah membutuhkan sebuah ciri dari suatu citra untuk dapat dikenali kemudian ciri tersebut akan dicocokkan dengan ciri citra lain. proses tersebut membutuhkan suatu metode ekstraksi ciri atau fitur dan klasifikasi.

Metode ekstraksi fitur dan jumlah neighbor yang melakukan voting mempengaruhi tingkat akurasi klasifikasi gambar. Proses pengenalan wajah dapat berbeda dari data objek yang digunakan.

Permasalahan yang sering terjadi dalam deteksi wajah adalah ketidakmampuan untuk mengenali wajah dengan tepat saat terjadi tumpang tindih, saat ada halangan pada citra wajah, atau ketidakjelasan yang lebih besar yang mengakibatkan proses deteksi wajah menjadi tidak akurat dengan gambar sebenarnya, terutama tidak akuratnya hasil antara wajah laki-laki dan perempuan. Oleh sebab itu, dibutuhkan proses metode klasifikasi untuk lebih meningkatkan hasil deteksi wajah.

Klasifikasi merupakan sebuah proses pengelompokan fitur citra pelatihan yang telah disimpan dengan fitur citra uji[8]. Salah satu bagian dari image recognition yang telah dikembangkan saat ini adalah pengenalan jenis kelamin (gender recognition). Kemiripan antara gender recognition dengan face recognition terletak pada proses ekstraksi fiturnya. Namun, sedikit berbeda pada proses pengklasifikasiannya[9]. Kesulitan dalam proses gender recognition terutama karena kekompleksan dari kondisi wajah, seperti posisi gambar, pencahayaan dan ekspresi gambar yang berbeda-beda yang memiliki dimensi serta reduksi tinggi sehingga harus melalui proses kompresi atau ekstraksi terlebih dahulu sebelum diolah datanya dengan metode klasifikasi.

Algoritma K-Nearest Neighbor (KNN) akan mengklasifikasikan citra uji ke dalam kelas dengan jumlah anggota terbanyak. Prinsip kerja dari algoritma KNN yaitu dengan mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (Neighbor)[10]. Tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan sample dari data training[11]. KNN menggunakan neighborhood classification sebagai nilai prediksi dari nilai instance yang baru berdasarkan atribut yang sudah dikehendaki terhadap masing-masing penelitian. Atribut yang sudah dipilih nantinya akan sebagai parameter untuk menentukan kelas[12].

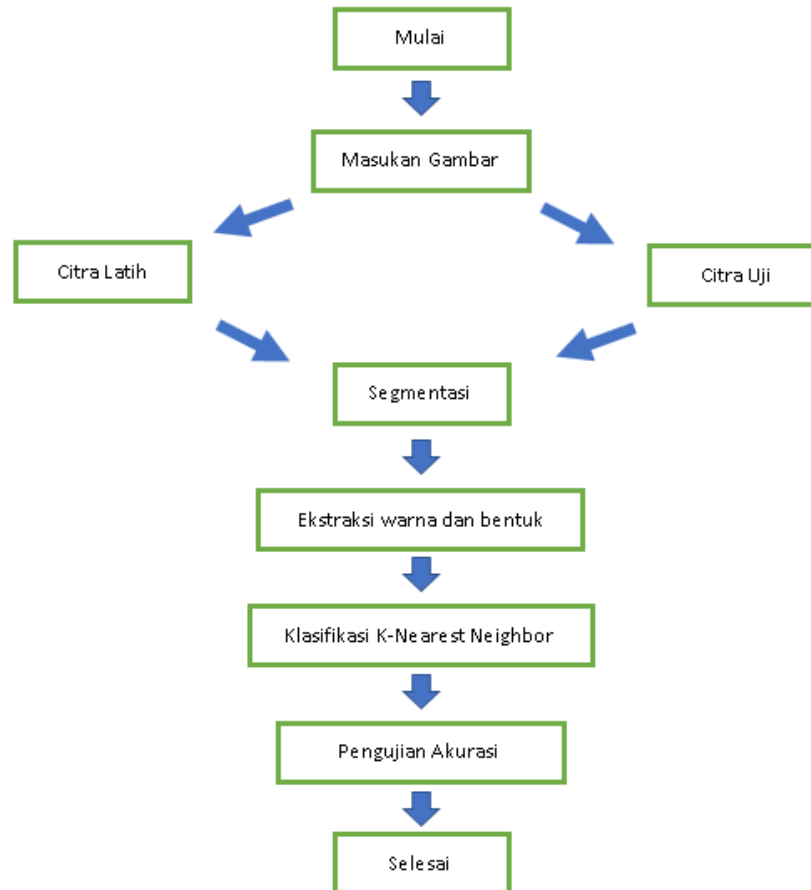
K-Nearest Neighbor termasuk salah satu algoritma paling sederhana yang digunakan dalam machine learning untuk regresi dan klasifikasi[13]. KNN mengikuti strategi “bird of a feather” dalam menentukan di mana data baru sebaiknya ditempatkan. Algoritma KNN mengasumsikan bahwa sesuatu yang mirip akan ada dalam jarak yang berdekatan atau bertetangga[14]. Artinya data-data yang cenderung serupa akan dekat satu sama lain. Kelebihan algoritma KNN adalah Untuk dataset berukuran besar, cost untuk menghitung jarak antara titik baru dan setiap titik yang ada sangat besar dan cenderung menurunkan kinerja algoritma, Algoritma KNN tidak bekerja dengan baik pada data berdimensi tinggi karena dengan jumlah dimensi yang besar, menjadi sulit bagi algoritma untuk menghitung jarak di setiap dimensi, Kita perlu melakukan penskalaan fitur (standarisasi dan normalisasi) sebelum menerapkan algoritma KNN ke kumpulan data apa pun[15]. Jika kita tidak melakukannya, KNN dapat menghasilkan prediksi yang salah.

Identifikasi masalah pada penelitian ini adalah belum adanya penelitian terkait mengenai pengenalan citra wajah menggunakan ekstraksi warna dan bentuk dengan penerapan metode klasifikasi dengan menggunakan fitur warna HSV dan RGB dengan menambahkan fitur bentuk area, eccentricity dan metric. Karena pada dasarnya wajah seseorang merupakan salah satu bagian penting pada seseorang untuk dapat menentukan jenis perbedaan jenis kelamin yang merupakan informasi yang paling penting secara biologis.

Tujuan dalam penelitian deteksi wajah ini adalah untuk meningkatkan akurasi deteksi agar dapat mengenali wajah dengan lebih tepat khususnya membedakan jenis kelamin dari wajah seseorang. Ini melibatkan pengembangan dan evaluasi algoritma atau model. Dari proses ini diharapkan mampu mengenali wajah dengan lebih baik dalam berbagai kondisi.

2. METODOLOGI PENELITIAN

Teknik pengenalan citra yang dilakukan dalam penelitian ini yaitu melakukan kombinasi metode morfologi untuk proses segmentasi ekstraksi citra dan metode klasifikasi K-Nearest Neighbor untuk memperkuat tingkat keakuratan data hasil pengolahan citra. Metode dalam penelitian ini dilakukan dalam dua fase, yaitu fase pelatihan dan fase pengujian. Dalam fase pelatihan, tahapan yang dilakukan bertujuan untuk mendapatkan model berdasarkan subset citra yang disebut citra latih. Ada beberapa tahapan dalam proses Analisa citra wajah berdasarkan warna dan bentuk, tahapan tersebut termuat dalam diagram alur seperti di bawah ini[12]:



Gambar 1. Metode Penelitian

Secara keseluruhan, tahap-tahap proses pengenalan citra yang dilakukan dalam penelitian ini adalah sebagai berikut:

- Langkah pertama adalah menyiapkan data set gambar yang akan dianalisa. Sumber dataset adalah dari Mendeley Data, dipublikasikan pada tahun 2020, Kontibutor dataset ini adalah I Nyoman Gede Arya Astawa, seorang peneliti dari Politeknik Negeri Bali. Dataset gambar yang digunakan terdiri dari 10 orang dan setiap orang memiliki 20 gambar wajah, sehingga informasi gambar total digunakan dalam penelitian ini adalah 200 gambar wajah. Gambar wajah yang diambil tersebut memiliki background yang kompleks sehingga untuk proses selanjutnya harus memisahkan background dan objek (foreground) dari gambar tersebut.
- Selanjutnya adalah memasukkan data gambar ke dalam 1 folder / tempat pengolahan data agar data gambar bisa dicari dan diolah dengan mudah
- Dari 200 gambar yang telah diambil tersebut akan dipisahkan menjadi dua bagian, yaitu data latih yang terdiri dari 150 gambar dan data tes yang terdiri dari 50 gambar.
- Dari setiap gambar latih dan tes tersebut akan dilakukan segmentasi untuk memisahkan objek (foreground) dan background menggunakan metode K-Means Clustering.
- Setelah data citra biner didapatkan dari hasil segmentasi, proses selanjutnya adalah melakukan seleksi warna dan bentuk dari data tersebut, dimana data yang diambil dari fitur warna adalah data rata-rata dari HSV dan data rata-rata dari RGB, serta fitur bentuk yang diambil adalah dari area, *eccentricity* dan *metric*.
- Langkah selanjutnya adalah penerapan metode *K-Nearest Neighbor* dari data fitur warna dan bentuk yang telah didapat.
- Setelah dilakukan proses klasifikasi menggunakan metode *K-Nearest Neighbor*, selanjutnya adalah melakukan pengujian akurasi dari metode yang digunakan menggunakan Confusion Matrix dan Area Under Curve

2.1 Metode K-Nearest Neighbor

Proses perhitungan pada metode klasifikasi data citra wajah menerapkan algoritma *K-Nearest Neighbor* yang menggunakan pembelajaran supervised learning dimana hasil dari data yang diuji diklasifikasikan berdasarkan keanggotaan terdekat yang terbanyak dari data uji[13]. Pada tahap uji, jarak dari data uji terhadap data latih dihitung dan sejumlah keanggotaan yang paling dekat diambil. Ada beberapa tahapan dalam proses perhitungan metode klasifikasi *K-Nearest Neighbor* termuat dalam diagram alur seperti di bawah ini:



Gambar 2. Alur Perhitungan K-Nearest Neighbor

Tahapan proses metode klasifikasi *K-Nearest Neighbor* yang dilakukan dalam penelitian ini adalah sebagai berikut:

- Setelah mendapat data wajah hasil ekstraksi fitur warna dan bentuk, langkah pertama dalam Analisa metode *K-Nearest Neighbor* adalah melakukan perhitungan jarak Euclidean dari data tes yang didapat dari hasil ekstraksi citra. Adapun rumus menghitung Euclidean adalah:

$$d = |x_i - y_i| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

keterangan:

x_i = Data Uji

y_i = Data Latih

i = Variabel Data

d = jarak

- Setelah mendapatkan hasil jarak Euclidean, langkah selanjutnya adalah mengurutkan hasil perhitungan jarak Euclidean dari jarak yang paling kecil
- Langkah selanjutnya adalah mengambil Data sebanyak nilai K . Dalam metode *K-Nearest Neighbor*, pendekatan sederhana dalam menentukan nilai k bisa dilakukan dengan rumus:

$$K = \sqrt{n} \quad (2)$$

dimana n merupakan banyaknya data yang dihitung. Nilai k yang tinggi akan mengurangi noise pada klasifikasi

- Langkah selanjutnya adalah mengumpulkan kategori y , dimana y merupakan data jarak euclidean terkecil yang paling banyak terkumpul.
- Langkah selanjutnya adalah penentuan hasil klasifikasi dari beberapa kategori y yang terkumpul, dimana data yang paling mayoritas merupakan hasil klasifikasi yang didapat.

2.2 Kajian Pustaka

Penelitian yang dilakukan oleh Alifa Puteri Bilqis Salsabila, Rika Desma Yunita dan Chaerur Rozikin yang membahas mengenai klasifikasi K-Nearest Neighbor pada produksi tanaman hias. Permasalahan yang ada berdasarkan produksi tanaman hias yang melonjak pada tahun 2019 menjadikan para pecinta tanaman terus bertambah dan untuk meminimalisir permasalahan ketidak tahuan terhadap jenis-jenis bunga yang ada. Penerapan algoritma K-Nearest Neighbor serta ekstraksi ciri terhadap warna dan tekstur sangat membantu dalam pengolahan citra untuk mengidentifikasi bunga menjadi lebih mudah dan mempersingkat waktu, dengan akurasi terbesar 71% menggunakan nilai K-7, identifikasi bunga berhasil dilakukan

Penelitian yang dilakukan oleh Dimas Rossiawan Hendra Putra, Fitri Marisa dan Indra Dharma Wijaya yang membahas mengenai identifikasi wajah menggunakan metode Naïve Bayes. Identifikasi Wajah merupakan bagian penting dalam teknik pengolahan citra digital untuk menentukan ukuran, lokasi dan nilai gambar. Deteksi wajah adalah langkah dalam sistem pengenalan wajah yang digunakan untuk identifikasi pribadi, sistem pemantauan, hukum pidana, interaksi manusia dan komputer, dan sebagainya pada. Penelitian ini menampilkan deteksi wajah menggunakan Naïve Bayes Classifier dengan warna merah, hijau dan biru (RGB) fitur model. Penelitian ini menggunakan citra dengan variasi ekspresi wajah dari data latih dan data pengujian. Langkah pertama adalah membuat model warna kulit wajah dengan fitur model RGB yang telah dilakukan proses normalisasi citra dengan penskalaan keabuan, citra biner dan pemfilteran citra menggunakan max filter, lalu cari warna RGB rata-rata kulit wajah, dilanjutkan dengan membangun Gaussian distribusi yang menunjukkan klasifikasi warna kulit dan diproses dengan metode Naïve Bayes Classifier (NBC) dalam menentukan hasil klasifikasi. Dari hasil pengujian dengan total 70 data uji dan a threshold 0,1

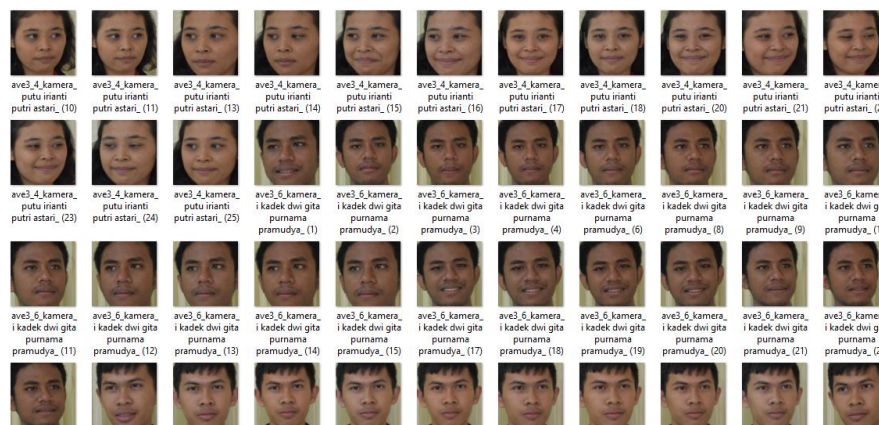
Penelitian yang dilakukan oleh I Kadek Surya Widiakumara, I Ketut Gede Darma Putra dan Kadek Suar Wibawa yang membahas mengenai identifikasi wajah. Identifikasi merupakan penentuan atau penetapan identitas seseorang dan proses mengidentifikasi adalah kegiatan dalam menentukan atau menetapkan identitas seseorang. Pengembangan teknologi identifikasi ini telah diterapkan pada macam-macam perangkat salah satunya pada smartphone berbasis Android. Kebanyakan dari pengembangan identifikasi berbasis Android masih menggunakan teknologi penyimpanan pada perangkat itu sendiri. Metode Eigenface digunakan untuk mengekstrak informasi yang relevan dari sebuah citra wajah, kemudian mengubahnya kedalam satu set kode yang paling efisien dan kode tersebut dibandingkan dengan kode dari citra wajah yang telah disimpan pada basis data. Aplikasi Identifikasi Wajah Berbasis Android ini dibangun dengan

menggunakan teknologi penyimpanan pada server (MySQL) dan juga menggunakan Meode Eigenface. Tingkat keberhasilan dari uji coba identifikasi wajah sebesar 68% dan tingkat salah pengenalan sebesar 32%, dari total uji coba sebanyak 25 kali identifikasi. Beberapa faktor penting yang mempengaruhi tingkat keberhasilan identifikasi yaitu posisi wajah dan intensitas cahaya saat melakukan pendaftaran

3. HASIL DAN PEMBAHASAN

3.1 Data Citra Collection

Data set yang digunakan terdiri dari 10 jenis spesies citra wajah dan masing-masing spesies memiliki 20 citra wajah sehingga total data citra yang digunakan sebagai dataset berjumlah 200 data. Dari seluruh dataset ini kemudian dibagi menjadi 2 bagian yaitu 80% akan digunakan sebagai data tes yaitu sebanyak 150 data citra dan 20% akan digunakan sebagai data latih yaitu sebanyak 50 data citra.



Gambar 3. Contoh Dataset

3.2 Ekstraksi warna dan bentuk

Selanjutnya adalah melakukan proses ekstraksi warna dan bentuk menggunakan metode morfologi yang bertujuan untuk menghilangkan noise pada gambar yang sudah di filter sehingga dapat memberikan hasil segmentasi yang lebih akurat[16]. Tahap akhir dari segmentasi ini adalah mendapatkan data citra biner dari hasil operasi morfologi citra.

Hasil dari ekstraksi citra, selanjutnya dibuat data citra training yang merupakan data hasil segmentasi menggunakan metode K-Means Clustering dan ekstraksi citra yang telah disederhanakan agar proses Analisa menggunakan metode K-Nearest Neighbor lebih terfokus. Data ini berisi fitur warna yang terdiri dari nilai Mean H, Mean S, Mean V, Mean Y, Mean CB, Mean CR dan fitur bentuk dari citra wajah yang terdiri dari area, eccentricity dan metric[17].

Setelah selesai proses ekstraksi warna dan bentuk menggunakan data tes, selanjutnya adalah fase pengujian, yaitu tahapan yang dilakukan bertujuan untuk mendapatkan model berdasarkan citra tes. Data yang diproses berjumlah 150 data citra. Proses yang dilakukan sama seperti yang diterapkan pada data tes, yaitu proses segmentasi hingga ekstraksi warna dan bentuk

Tahap akhir dari segmentasi ini adalah mendapatkan data citra biner dari hasil operasi morfologi citra yang ditampilkan pada tabel 1, berupa data Nama File, Label, R_mean, R_max, R_min, G_mean, G_max, G_min, B_mean, B_max, B_min, H_mean, H_max, H_min, S_mean, S_max, S_min, V_mean, V_max, V_min, area, perimeter, eccentricity

Nama File	Label	R_mean	R_max	R_min	G_mean	G_max	G_min	B_mean	B_max	B_min	H_mean	H_max	H_min	S_mean	S_max	S_min	V_mean	V_max	V_min	area	perimeter	eccentricity
ave3_1_kamera_i made dwiki satria wibawa_ (1)	Pria	163.91	188	55	158.09	184	52	125.21	159	35	0.1427	0.2045	0.0833	0.2416	0.4531	0.1319	0.6428	0.7373	0.2157	7370	553.97	0.9816
ave3_1_kamera_i made dwiki satria wibawa_ (10)	Pria	51.89	156	12	45.88	156	10	42.87	138	9	0.3019	0.9896	0.0000	0.1776	0.6333	0.0000	0.2053	0.6118	0.0471	8213	1269.30	0.9159
ave3_1_kamera_i made dwiki satria wibawa_ (11)	Pria	50.40	148	7	44.30	140	10	40.02	121	7	0.1647	0.9917	0.0000	0.1880	0.6563	0.0000	0.1991	0.5804	0.0471	8085	1201.68	0.9213
ave3_1_kamera_i made dwiki satria wibawa_ (12)	Pria	66.73	180	10	57.39	177	9	51.89	160	10	0.2021	0.9917	0.0000	0.2206	0.5952	0.0000	0.2629	0.7059	0.0392	9298	1218.31	0.8974
ave3_11_kamera_norasela_ (1)	Wanita	46.25	190	7	44.95	193	8	38.73	177	2	0.2069	0.9815	0.0000	0.1816	0.7500	0.0000	0.1846	0.7569	0.0314	3019	309.07	0.9448
ave3_11_kamera_norasela_ (10)	Wanita	144.30	178	35	140.71	177	36	103.36	146	20	0.1530	0.1869	0.1218	0.2881	0.4634	0.1350	0.5663	0.6980	0.1412	3979	516.43	0.9937
ave3_11_kamera_norasela_ (12)	Wanita	157.52	188	31	154.04	186	30	119.92	160	10	0.1530	0.1869	0.1282	0.2452	0.6774	0.1092	0.6182	0.7373	0.1216	4763	507.62	0.9951
ave3_11_kamera_norasela_ (13)	Wanita	160.31	189	37	156.52	186	36	121.87	166	17	0.1515	0.1961	0.1267	0.2447	0.5526	0.1027	0.6290	0.7412	0.1451	5074	527.97	0.9943
ave3_14_kamera_gusti ayu nyoman sita wahana	Wanita	135.73	163	39	130.58	158	34	90.63	118	14	0.1475	0.1667	0.1111	0.3368	0.6410	0.2303	0.5323	0.6392	0.1529	4299	522.28	0.9918
ave3_14_kamera_gusti ayu nyoman sita wahana	Wanita	134.27	162	41	130.98	158	40	92.99	123	20	0.1533	0.1825	0.1272	0.3109	0.5122	0.2065	0.5266	0.6353	0.1608	3661	517.08	0.9923
ave3_14_kamera_gusti ayu nyoman sita wahana	Wanita	133.22	163	51	130.60	159	48	93.56	124	30	0.1558	0.1825	0.1200	0.3019	0.4545	0.2051	0.5226	0.6392	0.2000	3856	514.46	0.9924
ave3_2_kamera_muhammad ilham maulana_ (10)	Pria	27.53	113	9	24.43	110	7	23.84	96	4	0.3664	0.9917	0.0000	0.1663	0.7333	0.0000	0.1100	0.4431	0.0353	6890	712.74	0.9180
ave3_2_kamera_muhammad ilham maulana_ (12)	Pria	27.71	180	10	23.80	181	8	22.92	165	6	0.4533	0.9917	0.0000	0.1968	0.5938	0.0000	0.1098	0.7098	0.0392	6436	698.83	0.9439
ave3_2_kamera_muhammad ilham maulana_ (13)	Pria	29.48	188	12	26.12	190	9	25.57	176	9	0.5066	0.9896	0.0000	0.1801	0.6071	0.0000	0.1178	0.7451	0.0471	6324	712.04	0.9447
ave3_2_kamera_muhammad ilham maulana_ (14)	Pria	162.62	188	31	161.96	191	31	137.14	174	21	0.1652	0.2500	0.0802	0.1654	0.4105	0.0749	0.6412	0.7490	0.1216	6956	575.59	0.9868

Gambar 4. Contoh data hasil segmentasi

Citra training merupakan data hasil segmentasi menggunakan metode K-Means Clustering dan ekstraksi citra yang telah disederhanakan agar proses Analisa menggunakan metode K-Nearest Neighbor lebih terfokus[18]. Data ini

berisi fitur warna yang terdiri dari nilai Mean H, Mean S, Mean V, Mean Y, Mean CB, Mean CR dan fitur bentuk dari citra wajah yang terdiri dari area, eccentricity dan metric. Pada tabel 2 dibawah ini merupakan data detil fitur warna hasil segmentasi yang telah disederhanakan, yang didapat digunakan untuk proses klasifikasi K-Nearest Neighbor.

Tabel 1. Contoh Data citra latih yang telah disederhanakan

Nama File	RGB Mean			HSV Mean			area	perimeter	eccentricity	Label
	R_mean	G_mean	B_mean	H_mean	S_mean	V_mean				
ave3_1_kamera_i made dwiki satria wibawa_ (15)	160.13	155.61	123.03	0.1475	0.2359	0.6281	8898	651.32	0.9759	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (16)	160.21	155.59	125.59	0.1463	0.2214	0.6288	7868	721.58	0.9781	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (17)	49.79	43.67	40.62	0.2403	0.1907	0.1979	8986	1473.63	0.9050	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (18)	161.26	157.80	126.54	0.1538	0.2284	0.6341	4649	562.37	0.9803	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (2)	145.48	139.56	103.24	0.1444	0.2931	0.5706	2383	264.38	0.9271	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (20)	35.75	31.11	28.09	0.3259	0.1897	0.1407	2288	315.46	0.9406	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (22)	48.25	42.17	39.46	0.2996	0.1942	0.1915	8801	1300.48	0.8996	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (23)	44.63	37.77	34.10	0.2981	0.2320	0.1764	7957	1164.56	0.9037	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (24)	45.66	39.85	37.22	0.3440	0.1880	0.1805	8726	1392.31	0.8928	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (25)	47.66	40.90	37.93	0.3388	0.2037	0.1880	8426	1391.99	0.9060	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (3)	178.49	175.65	153.73	0.1493	0.1402	0.7011	1290	376.55	0.9981	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (4)	46.49	39.48	36.42	0.2687	0.2192	0.1845	8492	1287.72	0.9037	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (6)	56.09	48.23	44.88	0.2558	0.2155	0.2222	4747	971.68	0.8771	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (7)	53.00	46.39	42.82	0.2250	0.1815	0.2095	9447	1240.01	0.8923	Pria
ave3_1_kamera_i made dwiki satria wibawa_ (9)	49.51	41.48	38.19	0.3580	0.2272	0.1955	8746	1174.79	0.9000	Pria
ave3_11_kamera_norasela_ (15)	160.48	156.91	125.85	0.1499	0.2205	0.6295	4515	547.72	0.9952	Wanita

3.3 Klasifikasi K-Nearest Neighbor

Variable yang digunakan dalam perhitungan K-Nearest Neighbor menggunakan fitur warna yang terdiri dari Mean H, Mean S, Mean V, Mean Y, Mean CB, Mean CR dan fitur bentuk yang terdiri dari eccentricity dan metric yang merupakan hasil ekstraksi citra wajah. Data Ekstraksi citra hasil data fase pelatihan dan fase pengujian yang berjumlah 150 untuk data tes dan 50 untuk data latih

Perhitungan metode K-Nearest Neighbor menggunakan data training. Dari data latih diketahui jumlah kasus ada 50, yang terdiri dari 5 data hasil ekstraksi citra dari 10 data wajah orang yang berbeda.

Setelah mendapat data latih dari hasil ekstraksi fitur warna dan bentuk, langkah selanjutnya dalam Analisa metode K-Nearest Neighbor adalah melakukan perhitungan jarak Euclidean dari data latih terhadap data uji. Data uji disini merupakan salah satu data yang diambil dari table data uji. Pada tabel 3 dibawah ini adalah data uji yang digunakan, dimana kategori yang digunakan dalam data uji ini adalah wanita:

Tabel 2. Data Uji

R_mean	G_mean	B_mean	H_mean	S_mean	V_mean	area	perimeter	eccentricity	Label
160,48	156,91	125,85	0,1499	0,2205	0,6295	4515	547,72	0,9952	Wanita

Dengan menggunakan rumus pengukuran jarak Euclidian, maka langkah perhitungan jarak data latih terhadap data uji, dan pada tabel 4 merupakan data pertama dari data citra latih hasil ekstraksi citra dimana kategori jenis kelamin dari data ini adalah pria dan nilai variabelnya adalah sebagai berikut:

Tabel 3. nilai variabel data latih pertama

R_mean	G_mean	B_mean	H_mean	S_mean	V_mean	area	perimeter	eccentricity
--------	--------	--------	--------	--------	--------	------	-----------	--------------

163,91	158,09	125,21	0,14	0,24	0,64	7370,00	553,97	0,98
--------	--------	--------	------	------	------	---------	--------	------

Maka perhitungan jarak Euclidean adalah

$$d1 = \sqrt{(163,91 - 160,48)^2 + (158,09 - 156,91)^2 + (125,21 - 125,85)^2 + (0,14 - 0,1499)^2 + (0,24 - 0,2205)^2 + (0,64 - 0,6295)^2 + (7370,0 - 4515,0)^2 + (553,97 - 547,72)^2 + (0,98 - 0,9952)^2} = 2894,07$$

Perhitungan jarak data latih terhadap data uji dilakukan seperti contoh diatas hingga semua data latih dihitung semua. Setelah mendapatkan hasil jarak Euclidean, langkah selanjutnya adalah mengurutkan hasil perhitungan jarak Euclidean dari jarak yang paling kecil atau nilai terkecil. Tabel 5 dibawah ini merupakan sebagian data jarak Euclidean yang telah dihitung berdasarkan data latih.

Tabel 4. Contoh Data latih dengan Jarak Euclidean

Label	RGB			HSV			POLA		Distance
	R_mean	G_mean	B_mean	H_mean	S_mean	V_mean	area	perimeter	
Pria	163,91	158,09	125,21	0,14	0,24	0,64	7370,00	553,97	0,98
Pria	51,89	45,88	42,87	0,30	0,18	0,21	8213,00	1269,30	0,92
Pria	50,40	44,30	40,02	0,16	0,19	0,20	8085,00	1201,68	0,92
Pria	66,73	57,39	51,89	0,20	0,22	0,26	9298,00	1218,31	0,90
Pria	162,32	158,60	128,34	0,15	0,21	0,64	7597,00	666,55	0,99
Wanita	46,25	44,95	38,73	0,21	0,18	0,18	3019,00	309,07	0,94
Wanita	144,30	140,71	103,36	0,15	0,29	0,57	3979,00	516,43	0,99
Wanita	157,52	154,04	119,92	0,15	0,25	0,62	4763,00	507,62	1,00
Wanita	160,31	156,52	121,87	0,15	0,24	0,63	5074,00	527,97	0,99
Wanita	159,75	156,40	122,48	0,15	0,24	0,63	4233,00	554,05	1,00
Wanita	134,43	132,36	95,32	0,16	0,30	0,53	3760,00	513,19	0,99
Wanita	89,96	80,22	52,48	0,12	0,42	0,35	1564,00	202,78	0,92
Wanita	135,73	130,58	90,63	0,15	0,34	0,53	4299,00	522,28	0,99
Wanita	134,27	130,98	92,99	0,15	0,31	0,53	3661,00	517,08	0,99
Wanita	133,22	130,60	93,56	0,16	0,30	0,52	3856,00	514,46	0,99
Wanita	153,27	149,41	114,14	0,15	0,27	0,60	9452,00	584,88	0,97
Wanita	49,44	40,64	36,30	0,17	0,26	0,19	13203,00	951,13	0,88
Wanita	142,58	134,68	94,69	0,14	0,35	0,56	5134,00	534,53	0,97
Wanita	155,43	147,20	110,17	0,14	0,30	0,61	9606,00	643,26	0,97
Wanita	155,31	149,12	113,37	0,14	0,28	0,61	9483,00	679,89	0,97

Langkah selanjutnya adalah mengambil Data sebanyak nilai K. Dalam metode K-Nearest Neighbor. Langkah awal adalah melakukan perhitungan dengan nilai k = 2 atau 2 tetangga terdekat dari data uji, dari table jarak Euclidean diatas, didapati hasil nilai terkecil yaitu dengan jarak Euclidean 253.706 dengan label wanita dan 322.0904 dengan label wanita.

Kemudian dihitung kembali dengan nilai k = 5 atau 5 tetangga terdekat data dari uji, dari table diatas, didapati beberapa hasil nilai terkecil yaitu dengan jarak Euclidean 253.706, 322.0904, 794.4171 dan 869.0095 dengan label wanita dan 692.0791 dengan label pria. Hasil ini sesuai dengan data uji yang digunakan.

Kemudian dihitung kembali dengan nilai k = 9 atau 9 tetangga terdekat data dari uji, dari table diatas, didapati beberapa hasil nilai terkecil yaitu dengan jarak Euclidean 253.706, 322.0904, 794.4171, 869.0095, 949.0768 dan 1516.024 dengan label wanita dan 692.0791, 1232.33 dan 1757.913 dengan label pria. Hasil ini sesuai dengan data uji yang digunakan.

Kemudian dihitung kembali dengan nilai k = 13 atau 13 tetangga terdekat data dari uji, dari table diatas, didapati beberapa hasil nilai terkecil yaitu dengan jarak Euclidean 253.706, 322.0904, 794.4171, 869.0095, 949.0768, 1516.024, 1767.105, 1794.236, 1856.115 dan 1948.785 dengan label wanita dan 692.0791, 1232.33 dan 1757.913 dengan label pria. Hasil ini sesuai dengan data uji yang digunakan.

Ketiga perhitungan nilai K diatas merupakan bagian dari pengumpulan kategori y, dimana y merupakan data jarak euclidean terkecil yang paling banyak terkumpul. Dari hasil analisa pengumpulan nilai y berdasarkan nilai K 1, 5, 9, dan 13, maka didapati hasil, mayoritas jarak Euclidean terkecil yaitu dari label wanita, dan hasil ini sesuai dengan data uji dimana data yang digunakan adalah data dari ekstraksi gambar wajah wanita, sehingga didapati hasil hasil pengujian metode klasifikasi K-Nearest Neighbor menggunakan ekstraksi citra wajah adalah akurat.

3.4 Pengujian metode klasifikasi

3.4.1 Confusion Matrix.

Keakuratan hasil klasifikasi dapat diukur dengan menggunakan confusion matrix[19]. Confusion matrix adalah media yang berguna untuk menganalisis seberapa baik classifier dapat mengenali tupel dari kelas yang berbeda[20]. Misalkan terdapat dua kelas, maka akan diistilahkan sebagai tupel positif dan tupel negative. True positif mengacu pada tupel positif yang diberi label dengan tepat oleh classifier, semetara true negative adalah tupel positif yang diberi label dengan

tepat oleh classifier. False positif adalah tupel negative yang diberi label tidak tepat, false negatif adalah tupel positif yang diberi label dengan tidak tepat.

Pengujian confusion matrix menggunakan aplikasi Rapidminer berdasarkan data latih yang sama dengan yang digunakan dalam perhitungan K-Nearest Neighbor. Pada aplikasi rapidminer, pengujian confusion matrix menggunakan teknik fold cross validation dengan pengajuan nilai k seperti data pada perhitungan sample yaitu 2, 5, 9 dan 13[21]. Selanjutnya seperti yang ditunjukkan dalam tabel 6, dengan menggunakan rapidminer, dilakukan proses validasi dengan menggunakan data latih, dan memberikan data Accuracy, Precision, Recall

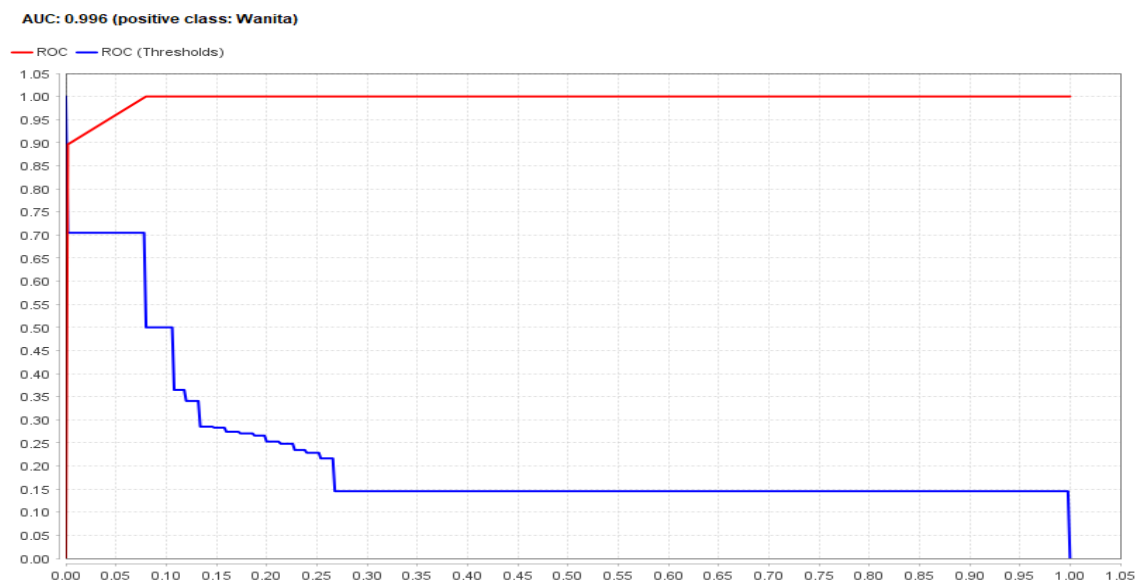
Tabel 5. Model Confusion Matrix

Validation	2	5	9	13
Accuracy	96%	83,3%	79,33%	79,33%
Precision	100%	85,71%	82,35%	82,35%
Recall	92%	80%	74,67%	74,67%

Dari hasil pengujian diatas yang ditunjukkan dalam tabel 6, maka didapat hasil data akurasi tertinggi menggunakan metode K-Nearest Neighbor berdasarkan data ekstraksi warna dan bentuk dari citra wajah dengan nilai K = 2 didapat nilai akurasi sebesar 96%.

3.4.2 Area Under Curve

Area Under Curve atau biasa disingkat AUC digunakan untuk mengukur perbedaan performansi dalam sebuah algoritma[22]. Pada penelitian ini pengujian nilai AUC masih menggunakan aplikasi rapidminer, menggunakan metode KNN dengan nilai k = 2 dikarenakan dalam pengujian validasi, nilai k = 2 merupakan nilai dengan akurasi tertinggi. dengan menggunakan data latih, didapat hasil kurva ROC seperti dibawah ini:



Gambar 5. Kurva ROC

Dari gambar kurva ROC diatas menunjukan hasil dari perhitungan metode k-nearest neighbor. Dalam gambar menunjukan nilai false positif rate yang ditujukan dengan garis vertikal memiliki rentang nilai 0 – 1, dan nilai true positive rate yang ditujukan dengan nilai horizontal memiliki rentang nilai 0 – 1[23][24]. Berdasarkan hasil kinerja pada gambar diatas, menunjukan kurva biru lebih bagus dibanding kurva merah, karena kurva biru berada jauh dari garis bujur (0,0). Untuk membandingkan nilai kinerja kurva merah dan biru dalam bentuk angka maka dapat dilakukan dengan membandingkan luas area kurva atau Area Under Curve (AUC)[25]. Pada gambar diatas, dapat disimpulkan luas area kurva berwarna biru lebih luas dibanding kurva berwarna merah, dengan nilai AUC sebesar 0.996.

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan pada citra wajah berdasarkan warna dan bentuk menggunakan metode K-Nearest Neighbor, dapat diambil kesimpulan bahwa metode ini termasuk dalam algoritma yang sangat baik untuk diaplikasikan pada klasifikasi citra wajah berdasarkan warna dan bentuk dengan nilai akurasi sebesar 96%, sehingga penentuan jenis kelamin berdasarkan objek wajah menggunakan data hasil ekstraksi warna dan bentuk serta menggunakan metode klasifikasi K-Nearest Neighbor sesuai dengan data citra sebenarnya. Dalam penelitian ini, peneliti hanya membatasi objek pada citra wajah untuk menentukan jenis kelamin, untuk lebih menguatkan hasil deteksi jenis kelamin, bisa ditambahkan objek lain seperti tinggi badan hingga postur tubuh, agar hasil yang didapat lebih optimal

REFERENCES

- [1] R. A. Asmara, B. S. Andjani, U. D. Rosiani, and P. Choirina, "Klasifikasi Jenis Kelamin Pada Citra Wajah Menggunakan Metode Naive Bayes," *J. Inform. Polinema*, vol. 4, no. 3, p. 212, 2018, doi: 10.33795/jip.v4i3.209.
- [2] D. R. H. Putra, F. Marisa, and I. D. Wijaya, "Identifikasi Wajah Berbasis Segmentasi Warna Kulit Wajah Menggunakan Naive Bayes Classifier," *J. Teknol. Inf.*, vol. 9, no. 2, pp. 99–106, 2018.
- [3] E. D. Sikumbang, "Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori," *J. Tek. Komput. AMIK BSI*, vol. Vol 4, No., no. September, pp. 1–4, 2018.
- [4] D. Hardiyanto and D. Anggun Sartika, "Optimalisasi Metode Deteksi Wajah berbasis Pengolahan Citra untuk Aplikasi Identifikasi Wajah pada Presensi Digital," *Setrum Sist. Kendali-Tenaga-elektronika-telekomunikasi-komputer*, vol. 7, no. 1, p. 107, 2018, doi: 10.36055/setrum.v7i1.3367.
- [5] A. Budi, S. Suma'inna, and H. Maulana, "Pengenalan Citra Wajah Sebagai Identifier Menggunakan Metode Principal Component Analysis (PCA)," *J. Tek. Inform.*, vol. 9, no. 2, pp. 166–175, 2018, doi: 10.15408/jti.v9i2.5608.
- [6] I. K. S. Widiakumara, I. K. G. D. Putra, and K. S. Wibawa, "Aplikasi Identifikasi Wajah Berbasis Android," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 8, no. 3, p. 200, 2017, doi: 10.24843/lkjiti.2017.v08.i03.p06.
- [7] N. Yelliy N, "Pengolahan Citra Digital Perbandingan Metode Histogram Equalization Dan Spesification Pada Citra Abu-Abu," *J-Icon*, vol. 7, no. 1, pp. 87–95, 2019.
- [8] R. Andrian, S. Anwar, M. A. Muhammad, and A. Junaidi, "Identifikasi Kupu-Kupu Menggunakan Ekstraksi Fitur Deteksi Tepi (Edge Detection) dan Klasifikasi K-Nearest Neighbor (KNN)," *J. Tek. Inform. dan Sist. Inf.*, vol. 5, no. 2, pp. 234–243, 2019, doi: 10.28932/jutisi.v5i2.1744.
- [9] D. P. Utomo, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. April, pp. 437–444, 2020, doi: 10.30865/mib.v4i2.2080.
- [10] N. Wijaya and A. Ridwan, "Klasifikasi Jenis Buah Apel Dengan Metode K-Nearest Neighbors," *Sisfokom*, vol. 08, no. 1, pp. 74–78, 2019.
- [11] H. Muchtar and F. Said, "Sistem Identifikasi Plat Nomor Kendaraan Menggunakan Metode Robert Filter dan Framing Image Berbasis Pengolahan Citra Digital," *Resist. (elektronika kEndali Telekomun. tenaga List. kOmputeR)*, vol. 2, no. 2, p. 105, 2019, doi: 10.24853/resistor.2.2.105-112.
- [12] T. Y. Prahudaya and A. Harjoko, "Metode Klasifikasi Mutu Jambu Biji Menggunakan Knn Berdasarkan Fitur Warna Dan Tekstur," *Jurnal Teknosains*, vol. 6, no. 2, p. 113, 2017, doi: 10.22146/teknosains.26972.
- [13] Johan Wahyudi and Ihdahubbi Maulida, "Pengenalan Pola Citra Kain Tradisional Menggunakan Glcm Dan Knn," *J. Teknol. Inf. Univ. Lambung Mangkurat*, vol. 4, no. 2, pp. 43–48, 2019, doi: 10.20527/jtiulm.v4i2.37.
- [14] A. P. W. Riri Nada Devita, Heru Wahyu Herwanto, "Perbandingan kinerja metode Naive Bayes dan KNN untuk klasifikasi artikel berbahasa indonesia," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 427–434, 2018, doi: 10.25126/jtiik.201854773.
- [15] R. Rahmadiano, E. Mulyanto, and T. Sutojo, "Implementasi Pengolahan Citra dan Klasifikasi K-Nearest Neighbor untuk Mendeteksi Kualitas Telur Ayam," *J. VOI (Voice Informatics)*, vol. 8, no. 1, pp. 45–54, 2019.
- [16] B. Salsabila, Alifa, Puteri, D. Yunita, Rika, and C. Rozikin, "Identifikasi Citra Jenis Bunga menggunakan Algoritma KNN dengan Ekstrasi Warna HSV dan Tekstur GLCM," *Technomedia J.*, vol. 6, no. 1, pp. 124–137, 2021, doi: 10.33050/tmj.v6i1.1667.
- [17] D. P. and A. B. S. Pamungkas, "IMPLEMENTASI EKSTRASI FITUR DAN K-NEAREST NEIGHTBOR UNTUK IDENTIFIKASI WAJAH PERSONAL," *Https://Medium.Com/*, vol. 3, no. 2, pp. 187–193, 2018, [Online]. Available: <https://medium.com/@arifwicaksanaa/pengertian-use-case-a7e576e1b6bf>.
- [18] N. Nafiah, "Klasifikasi Kematangan Buah Mangga Berdasarkan Citra HSV dengan KNN," *J. Elektron. List. dan Teknol. Inf. Terap.*, vol. 1, no. 2, pp. 1–4, 2019, [Online]. Available: <https://ojs.politeknikjambi.ac.id/elti>.
- [19] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.
- [20] T. Imandasari, E. Irawan, A. P. Windarto, and A. Wanto, "Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 750, 2019, doi: 10.30645/senaris.v1i0.81.
- [21] R. Evan Purnama Ramdan, Inti Mulyo Arti, "Identifikasi Dan Uji Virulensi Penyakit Antraknosa Pada Pasca panen Buah Cabai," *J. Pertan. Presisi (Journal Precis. Agric.)*, vol. 3, no. 1, pp. 67–76, 2019, [Online]. Available: <https://ejournal.gunadarma.ac.id/index.php/jpp/article/view/1976>.
- [22] A. Ciputra, D. R. I. M. Setiadi, E. H. Rachmawanto, and A. Susanto, "Klasifikasi Tingkat Kematangan Buah Apel Manalagi Dengan Algoritma Naive Bayes Dan Ekstraksi Fitur Citra Digital," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 465–472, 2018, doi: 10.24176/simet.v9i1.2000.
- [23] Maulana Fansyuri and O. Hariansyah, "Pengenalan Objek Bunga dengan Ekstraksi Fitur Warna dan Bentuk Menggunakan Metode Morfologi dan Naïve Bayes," *J. Sist. dan Inform.*, vol. 15, no. 1, pp. 70–80, 2020, doi: 10.30864/jsi.v15i1.338.
- [24] M. H. Rifqo and A. Wijaya, "Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit," *J. Pseudocode*, vol. 4, no. 2, pp. 120–128, 2017, doi: 10.33369/pseudocode.4.2.120-128.
- [25] R. A. Syawalia, S. Rasyad, and D. A. Pratama, "Implementasi Fuzzy Logic pada Sistem Sortir Otomatis Alat Penghitung Jumlah Buah Apel," *J. Tek. Elektro Dan Vokasional*, vol. 06, no. 02, pp. 421–432, 2020.