

IDENTIFIKASI VARIETAS JAGUNG MUTIARA BERDASARKAN DATA CITRA DIGITAL MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR*

Vicky Chandra¹, Jajam Haerul Jaman², Garno³

^{1,2,3}Teknik Informatika, Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang
E-mail: ¹vicky.16221@student.unsika.ac.id, ²jajam.haeruljaman@staff.unsika.ac.id,
³garno@staff.unsika.ac.id

Abstrak

Jagung adalah salah satu tanaman pangan penghasil karbohidrat yang penting di Indonesia selain padi dan gandum. Di Indonesia sendiri setiap tahun dapat menghasilkan jumlah jagung yang sangat banyak. Terbukti pada tahun 2017 Indonesia dapat memproduksi jagung sebanyak 22.59 juta ton. Jumlah ini sangat berarti bagi perekonomian di Indonesia. Salah satu jagung yang paling banyak ditanam di Indonesia adalah jagung mutiara. Namun, cukup sulit untuk membedakan setiap varietas jagung karena bentuk dan warna cenderung sama. Salah satu untuk mengetahui suatu varietas jagung, ialah dengan melakukan klasifikasi varietas jagung dengan memanfaatkan data hasil pengolahan citra setiap jagung. Salah satu metode klasifikasi yang digunakan ialah K-NN. Setiap citra jagung diambil nilai area, parameter, width, length, metric dan eccentricity menggunakan aplikasi Matlab guna mengetahui bentuk dari suatu jagung, lalu dikumpulkan sebanyak 250 data. Data yang telah dikumpulkan dibagi berdasarkan rasio perbandingan 70% data latih 30% data uji dan menerapkan nilai k yaitu 3, 5 dan 7. Proses klasifikasi menggunakan aplikasi Rstudio berbasis cloud. Hasil penelitian yang didapat, diketahui bahwa pengujian pada $k = 3$ mendapatkan nilai akurasi tertinggi dibandingkan nilai k yang lain. Akurasi yang didapat yakni sebesar 93.24% diikuti recall 91.89% precision 94.44% specificity 94.59% error rate 6.76% dan f-measure 0.9315.

Kata Kunci: Jagung, K-Nearest Neighbor, Pengolahan Citra Digital, Ekstrasi Ciri, Rstudio

Abstract

Corn is one of the most important carbohydrate-producing food crops in Indonesia besides rice and wheat. Every year Indonesia can produce a very large amount of corn. It is proven that in 2017 Indonesia can produce 22.59 million tons of corn. This number is very meaningful for the economy in Indonesia. One of the most widely grown corns in Indonesia is pearl corn. However, it is quite difficult to distinguish each variety of corn because the shape and color tend to be the same. One way to find out a corn variety is by utilizing the image processing data for each corn. One of the classification methods used is K-NN. Each corn image is taken the value of area, parameter, width, length, metric and eccentricity using the Matlab application to determine the shape of a corn, then 250 data are collected. The data that has been collected is divided based on a ratio of 70% training data 30% test data and applying k values, namely 3, 5 and 7. The classification process uses the cloud-based Rstudio application. The results obtained, it is known that the test at $k = 3$ gets the highest accuracy value compared to other k values. The accuracy obtained is 93.24% followed by recall 91.89% precision 94.44% specificity 94.59% error rate 6.76% and f-measure 0.9315.

Keywords: Corn, K-Nearest Neighbor, Digital Image Processing, Feature Extraction, R-Studio

1. PENDAHULUAN

Sektor pertanian merupakan sektor penyumbang PDB yang cukup besar meskipun perannya semakin menurun karena pertumbuhan di sektor non pertanian pertumbuhannya relatif lebih cepat. Pada tahun 2015, sumbangan sektor pertanian dalam arti sempit (subsektor tanaman pangan, hortikultura, perkebunan dan peternakan) terhadap PDB adalah 10,27%. Pada tahun 2019 sumbangan sektor pertanian terhadap PDB turun menjadi 9,41%. Menurunnya sumbangan sektor pertanian dalam arti sempit menunjukkan adanya transformasi perekonomian nasional, yang awalnya didominasi oleh hasil produk primer pertanian dalam arti sempit bergeser ke sektor lainnya [1].

Salah satu produk pertanian yang mengambil peran penting dalam pembangunan sektor pertanian adalah jagung. Pada tahun 2017 produksi jagung mencapai angka 28,92 juta ton pipilan kering atau naik 47,48% jika dibandingkan produksi tahun 2015, yang hanya sebesar 19,61 juta ton pipilan kering. Dari data tersebut kita telah mengetahui bahwa produksi jagung di Indonesia mengalami kenaikan pada jumlah produksi, artinya jagung masih menjadi salah satu pangan pokok yang banyak dikonsumsi oleh masyarakat Indonesia. Rata – rata tingkat konsumsi jagung pada tahun 2019 dengan presentase sebesar 4% per tahun [2].

Jagung mutiara adalah salah satu jenis jagung yang memiliki banyak khasiat untuk kesehatan. Jagung mutiara memiliki bulir yang tebal dan padat serta bentuk biji yang bulat. Jagung mutiara adalah salah satu jagung yang banyak ditanam oleh masyarakat Indonesia. Proses budidaya jagung memegang peranan sangat penting, dimana untuk mendapatkan hasil yang baik tentunya proses budidaya pun harus dilakukan dengan baik. Di antara tahapan budidaya jagung adalah pengklasifikasian varietas. Varietas tersebut dapat dibedakan berdasarkan morfologinya. Diantara varietas tersebut terdapat varietas unggul yang mempunyai beberapa keistimewaan dibandingkan dengan varietas lainnya.

Namun demikian, kendala besar yang dihadapi saat ini adalah masih kurangnya pengetahuan mengenai cara pengklasifikasian varietas dalam budidaya jagung karena masing – masing varietas.

Pengolahan citra digital merupakan teknik yang digunakan untuk merubah suatu bentuk citra menjadi bentuk lain yang memiliki tujuan untuk mempermudah manusia dalam mengidentifikasi suatu objek pada citra [3]. Sebelum melakukan proses pengolahan citra digital, maka perlu dilakukan proses akuisisi citra terhadap objek atau biji jagung mutiara yang akan diidentifikasi untuk mendapatkan data citra biji jagung mutiara. Akuisisi citra adalah tahap mengambil, mengumpulkan, dan menyiapkan data citra menggunakan alat pengambilan citra dan memprosesnya sehingga menghasilkan data yang dikehendaki [4]. Proses selanjutnya adalah mendapatkan ciri yang membedakan objek pada citra dengan objek lainnya. Proses tersebut yaitu melakukan ekstraksi ciri terhadap data citra jagung mutiara yang sebelumnya telah mengalami proses akuisisi citra. Hasil dari ekstraksi ciri yaitu nilai- nilai khusus yang mencirikan objek pada citra atau yang membedakannya dengan objek lain.

Oleh sebab itu digunakanlah teknik klasifikasi karena klasifikasi memiliki tugas yaitu memberikan label kelas pada sekumpulan *dataset*. Klasifikasi merupakan salah satu teknik dalam data mining yang mengelompokan data kedalam kelas tertentu berdasarkan pola dan atribut yang terdapat pada data latih yang telah dikelompokan sebelumnya [5]. Salah satu algoritma yang dapat digunakan untuk klasifikasi adalah algoritma *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* merupakan sebuah metode yang dapat mengklasifikasikan objek pada data berdasarkan jarak terdekat antara data pembelajaran atau data latih dengan objek yang sedang diteliti [6]. Berdasarkan prinsip kerja dari algoritma *K-Nearest Neighbor* yang menemukan objek berdasarkan jarak dengan data latih menggunakan nilai *k* tertentu, maka penelitian ini mencoba menerapkan

algoritma tersebut untuk mengklasifikasi varietas jagung mutiara. Sehingga hasilnya adalah kelas pada objek yang memiliki kriteria tertentu berdasarkan jarak dengan nilai k terdekat.

Penelitian pertama mengenai pengolahan citra digital dan klasifikasi menggunakan algoritma *K-Nearest Neighbor* pernah dilakukan oleh Rusli dan Nasir. Penelitian tersebut bertujuan untuk mengetahui jenis mangga berdasarkan bentuk daunnya. Untuk mengetahui pola daun dapat dibedakan berdasarkan bentuk dan tepi daun. Pada penelitian tersebut untuk menghitung jaraknya menggunakan metode *Euclidean Distance* dan memiliki nilai akurasi tertinggi dengan nilai $k = 1$ [7].

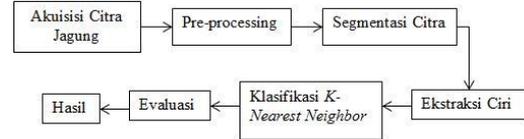
Penelitian lainnya pernah dilakukan oleh Hasson dan Jalil yang menggunakan algoritma *K-Nearest Neighbor*. Tujuan dari penelitian tersebut adalah untuk mengklasifikasikan jenis beras Anber Iraq menggunakan metode *image processing*. Pada penelitian tersebut menggunakan metode *Euclidean Distance* untuk menghitung jaraknya dan memiliki nilai akurasi sebesar 83% dengan nilai $k = 1$ [8].

Berdasarkan permasalahan, data, dan studi literatur yang telah diuraikan maka salah satu untuk mengatasi hal tersebut, penelitian ini melakukan klasifikasi dengan metode K-NN menggunakan sejumlah data pengolahan citra jagung yang telah dikumpulkan sebelumnya. Dalam penelitian ini ditentukan K yang digunakan adalah 3, 5 dan 7 dengan pembagian data untuk data latih dan data uji dengan rasio pembagian yakni 70% untuk data latih dan 30% data uji.

1. METODE PENELITIAN

Penelitian ini dilakukan menggunakan data hasil pengolahan citra pada jagung. Data diperoleh dari citra jagung yang kemudian citra tersebut diambil nilai khusus yang berupa *area*, *parameter*, *length*, *width*, *eccentricity*, dan *metric*. Kemudian data tersebut diolah menggunakan teknik data mining yakni menggunakan algoritma *K-Nearest Neighbor* dengan bantuan tools Rstudio menggunakan bahasa pemrograman R.

Tahap terakhir melakukan evaluasi berdasarkan parameter *accuracy*, *recall*, *precision*, *specificity*, *error rate* dan *f-measure*. Berikut metodologi penelitian yang digunakan:



Gambar 1. Metodologi Penelitian

2.1 Akuisisi Citra

Akuisisi citra adalah proses mengambil dan mengumpulkan citra menggunakan alat pengambilan citra. Proses akuisisi citra dilakukan oleh penulis sendiri menggunakan kamera *handphone* Xiaomi dengan resolusi kamera sebesar 16 MP dengan jarak antara objek dengan kamera sebesar 10 cm. Akuisisi citra dilakukan untuk mengumpulkan data citra jagung mutiara yang dibutuhkan dalam penelitian. Proses akuisisi citra dilakukan terhadap setiap butir jagung. Sebanyak 125 butir varietas jagung mutiara dilakukan proses akuisisi citra untuk menghasilkan citra jagung mutiara. Proses akuisisi citra juga dilakukan terhadap varietas jagung lain yang diambil secara acak sebanyak 125 butir. Jadi total butir jagung yang dilakukan proses akuisisi citra sebanyak 250 butir yang terdiri dari citra varietas jagung mutiara dan citra varietas jagung lain.

2.2 Pre-Processing

Pre-processing merupakan proses persiapan data hingga data siap untuk dilakukan proses klasifikasi. Tahap ini juga disebut sebagai tahap pembersihan data. Pada tahap *pre-processing* yang dilakukan pada penelitian ini yaitu melakukan perubahan ukuran, warna, dan kualitas citra. Adapun tahap-tahap yang dilakukan pada proses *pre-processing* hingga data siap dilakukan proses klasifikasi yaitu :

1. Melakukan *cropping* citra. Tahap ini dilakukan untuk menghilangkan bagian-bagian yang tidak terpakai pada citra jagung.
2. Mengubah citra RGB menjadi citra biner. Dengan mengubah citra RGB

menjadi citra biner, hal ini akan memudahkan dalam pengolahan citra ke tahap selanjutnya.

3. Melakukan perbaikan kualitas citra. Perbaikan kualitas citra dilakukan untuk meningkatkan kualitas pada citra. Tahapan ini tidak harus dilakukan jika citra telah memiliki kualitas yang baik.

2.3 Segmentasi Citra

Segmentasi citra dilakukan sebelum tahap ekstraksi ciri atau pengenalan pola. Segmentasi citra dilakukan untuk memisahkan objek pada citra dengan *background*. Dengan kata lain segmentasi citra digunakan untuk memisahkan objek yang memiliki informasi yang dibutuhkan dengan objek lain yang tidak dibutuhkan. Proses ini dilakukan untuk memudahkan tahap selanjutnya yaitu melakukan ekstraksi ciri. Metode segmentasi citra yang digunakan pada penelitian ini adalah metode *thresholding*. Metode ini bekerja berdasarkan tingkat kecerahan. Cara kerjanya adalah pada bagian citra jagung yang cenderung gelap akan dibuat semakin gelap. Sedangkan pada bagaian citra jagung yang cenderung terang akan dibuat semakin terang. Hasil dari segmentasi dengan menggunakan metode *thresholding* adalah citra biner yang memiliki nilai intensitas pixel 1 dan 0 [9].

2.4 Ekstraksi Ciri

Ekstaksi ciri merupakan suatu proses untuk mendapatkan fitur atau ciri-ciri dari suatu citra atau gambar. Ciri-ciri atau karakteristik tersebut yang akan digunakan sebagai parameter untuk menggambarkan suatu objek, dan parameter itu juga yang akan dijadikan sebagai data masukkan dalam proses klasifikasi [10, 11]. Tujuan dari ekstraksi ciri yaitu untuk mendapatkan nilai-nilai khusus pada citra jagung berdasarkan parameter-parameter yang mencirikan citra tersebut. Citra asli tanpa *background* yang diperoleh dari segmentasi citra selanjutnya dilakukan proses perhitungan untuk mendapatkan nilai *perimeter*, *area*, *eccentricity*, *width*, dan *length*. Selanjutnya nilai-nilai yang telah

didapatkan tersebut dianalisis menggunakan teknik klasifikasi.

2.5 Klasifikasi *K-Nearest Neighbor*

K-Nearest Neighbor merupakan algoritma data mining dalam memproses klasifikasi klasik yang sederhana. Metode ini juga sering disebut sebagai *Instance Based Learning* karena, K-NN melakukan klasifikasi berdasarkan jarak antara suatu objek dengan objek yang lain. Prinsip dari K-NN adalah mencari jarak terdekat dari dua objek dengan k tetangga terdekatnya dalam data pelatihan [12, 13]. Pada tahap klasifikasi ini nilai *perimeter*, *area*, *eccentricity*, *width*, dan *length* yang telah didapatkan dari hasil ekstraksi ciri terhadap 250 citra jagung, kemudian digabungkan dalam bentuk tabel untuk mempermudah melakukan proses klasifikasi. Proses pembuatan tabel menggunakan *Ms. Excel* dan menghasilkan 250 *record* atau baris pada tabel yang memiliki 6 atribut yaitu *Perimeter*, *Area*, *Eccentricity*, *Width*, *Length*, dan Label Kelas.

Sebelum memasuki tahap klasifikasi harus dilakukan pembagian data *training* dan data *testing*. Pembagian data *training* dan data *testing* pada penelitian ini menggunakan perbandingan 7:3, dimana 70% merupakan data *training* dan 30% merupakan data *testing* dari total 250 *record* data. Jumlah *record* pada data *training* sebanyak 175 sedangkan pada data *testing* sebanyak 75 *record* data. Tahapan klasifikasi menggunakan algoritma *K-Nearest Neighbor* pada penelitian ini adalah sebagai berikut:

1. Menghitung Jarak Antara Data *Training* dan Data *Testing*. Perhitungan jarak pada penelitian ini menggunakan rumus *Euclidean Distance*. Nilai *perimeter*, *area*, *eccentricity*, *width*, dan *length* yang terdapat pada data *testing* dihitung jaraknya terhadap nilai *perimeter*, *area*, *eccentricity*, *width*, dan *length* yang terdapat pada data *training*. Jarak dihitung dari setiap *record* pada data *testing* terhadap semua *record* pada data *training*.

2. Mengurutkan Data Hasil Perhitungan. Jarak yang telah didapatkan berdasarkan perhitungan menggunakan rumus *Euclidean Distance* diurutkan berdasarkan nilai terkecil hingga nilai terbesar. Jarak yang telah didapatkan berdasarkan perhitungan menggunakan rumus *Euclidean Distance* diurutkan berdasarkan nilai terkecil hingga nilai terbesar.
3. Menentukan Nilai k. Penentuan nilai k atau jumlah tetangga terdekat dalam klasifikasi menggunakan algoritma *K-Nearest Neighbor* tidak memiliki aturan yang baku. Pada penelitian ini ditetapkan nilai tetangga terdekat atau k yaitu sebesar 3, 5, dan 7.
4. Menentukan Kelas Label Berdasarkan Mayoritas k Tetangga. Penentuan label kelas pada data *testing* ditentukan berdasarkan jumlah mayoritas k tetangga terdekat pada data *training*. Jika mayoritas kelas label adalah “Ya”, artinya objek tersebut merupakan varietas jagung mutiara. Akan tetapi jika mayoritas kelas label adalah “Tidak”, maka artinya objek tersebut bukan merupakan varietas jagung mutiara.

Adapun rumus dari *Euclidian distance* sebagai berikut:

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

Keterangan :

d = jarak antara nilai x dengan nilai y
x = sampel data atau data training
y = data uji atau data testing
i = nilai atribut pada data
n = jumlah data

2.6 Evaluasi

Peran evaluasi sangat penting dalam *data mining*, karena evaluasi digunakan untuk mengevaluasi kualitas dan efektivitas dari model atau metode-metode data mining sebelum diterapkan. Penerapan model dari *data mining* termasuk ke dalam investasi yang dilakukan oleh perusahaan, jika model yang digunakan tidak valid maka waktu dan uang perusahaan akan terbuang sia-sia [14].

Tahap evaluasi dilakukan untuk mengetahui seberapa baik algoritma *K-Nearest Neighbor* dalam melakukan klasifikasi terhadap varietas jagung mutiara. Ukuran dan parameter yang digunakan untuk tahap evaluasi mengacu kepada nilai yang didapatkan pada *confusion matrix*. Dan selanjutnya akan dilakukan proses perhitungan untuk mencari nilai *Accuracy*, *Precision*, *Recall*, *Specificity*, *Error Rate* dan *F-Measure*.

2.6.1 Confusion Matrix

Untuk mengetahui seberapa baik prediksi yang telah dibuat dapat menggunakan tools *confusion matrix*. TP dan TN menjelaskan bahwa kapan prediksi yang dibangun melakukan sesuatu dengan benar, sementara FP dan FN menjelaskan kapan prediksi yang dibangun melakukan sesuatu kesalahan [15].

Tabel 1. Confusion Matrix

		Actual class		
Predictiton	Tidak	Tidak (TP)	Ya (FP)	
	Ya		(FN)	(TN)

Keterangan:

True Positive (TP) : TP merujuk pada jumlah benar positif yang diberi label dengan benar saat klasifikasi.
True Negative (TN) : TN merujuk pada jumlah benar negatif yang diberi label dengan benar saat klasifikasi.
False Positive (FP) : FP merujuk pada jumlah salah negatif yang diberi label yang salah dan diklaim sebagai positif.
False Negative (FN) : FN merujuk pada jumlah salah positif yang diberi label yang salah dan diklaim sebagai negatif.

Selain itu dalam evaluasi metode klasifikasi terdapat beberapa perhitungan seperti menghitung *accuracy*, *error rate*, *precision*, *recall*, dan *F-Measure* dan lain-lain [16].

1. Accuracy

Akurasi merupakan rasio prediksi benar (TP dan TN) dengan keseluruhan data. *Accuracy* didefinisikan sebagai tingkat hubungan antara nilai prediksi dengan nilai aktual. Akurasi ini sangat bagus apabila dijadikan sebagai tolak ukur pada

performansi dari algoritma yang digunakan, apabila jumlah data FN dan FP sangat mendekati [13]. Berikut ini adalah rumus mencari akurasi dari tabel *confusion matrix*:

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (2)$$

2. Precision

Setelah mencari nilai akurasi, selanjutnya mencari nilai *precision* yang merupakan rasio prediksi benar positif yang dibandingkan dengan keseluruhan hasil yang diprediksi positif dengan artian bahwa *precision* adalah ketepatan antara hasil yang diberikan oleh sistem dengan harapan informasi pengguna yang diinginkan. Di mana *precision* memiliki rumus yaitu:

$$Precision = \frac{(TP)}{(TP+FP)} \quad (3)$$

3. Recall

Recall merupakan rasio prediksi benar positif yang dibandingkan dengan keseluruhan data yang benar positif di mana *recall* adalah tingkat keberhasilan sistem menemukan kembali sebuah informasi. *Recall* memiliki rumus yakni:

$$Recall = \frac{(TP)}{(TP+FN)} \quad (4)$$

4. Specificity

Specificity merupakan nilai keberhasilan memprediksi negatif terhadap keseluruhan data negatif. *Specificity* memiliki rumus:

$$Specificity = \frac{(TN)}{(TN+FP)} \quad (5)$$

5. Error Rate

Error Rate digunakan untuk menghitung rasio kesalahan dalam memprediksi dari semua total data yang dievaluasi. *Error Rate* memiliki rumus:

$$Error Rate = \frac{(FP+FN)}{(TP+FP+FN+TN)} \quad (6)$$

6. F-measure

F-measure digunakan untuk mengetahui hubungan timbal balik antara *precision* dan *recall*. *F-measure* memiliki rumus:

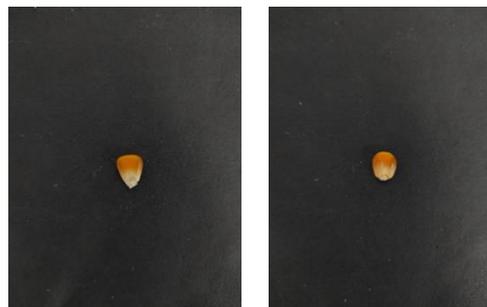
$$F\text{-measure} = 2x \frac{(Precision \times Recall)}{(Precision+Recall)} \quad (7)$$

3. HASIL DAN PEMBAHASAN

3.1 Akuisisi Citra

Tahap ini melakukan pengambilan citra jagung mutiara. Pengambilan citra dilakukan menggunakan kamera dengan

kemampuan 16 megapixel. Pengambilan gambar dilakukan dengan jarak sejauh 10cm dari objek dan diberikan background berwarna hitam. Varietas jagung yang dilakukan proses akuisisi citra adalah jagung varietas mutiara sejumlah 125 butir. Selain itu juga dilakukan proses akuisisi citra terhadap 125 varietas jagung lain sebagai data acuan yang membedakan dengan varietas jagung mutiara. Adapun varietas jagung lain yang digunakan adalah Ketan Putih, Manis, Brondong, dan Kristal. Untuk varietas jagung ketan putih, brondong dan kristal masing-masing berjumlah 30 butir jagung sedangkan untuk jagung manis sebanyak 35 butir. Sehingga total butir jagung yang dilakukan proses akuisisi citra sebanyak 250 butir.



Gambar 2. Hasil Akuisisi Citra Jagung

3.2 Preprocessing

Preprocessing pada citra digital dilakukan untuk meningkatkan kualitas citra agar layak diimplementasikan pada tahap selanjutnya. Terdapat dua tahapan yang dilakukan pada pre-processing, yaitu:

3.2.1 Mengubah Ukuran Citra

Ukuran semula pada citra yaitu 1280x1600, dan akan melalui proses perubahan menjadi 300x300. Proses mengubah ukuran pada citra dilakukan untuk menghilangkan bagian yang tidak terpakai. Selain itu proses ini dilakukan agar ukuran citra lebih seragam. Mengubah ukuran citra dengan proses *cropping* ini dilakukan menggunakan tools Matlab R2019a. Berikut adalah hasil dari mengubah ukuran citra jagung.



Gambar 3. Hasil Mengubah Ukuran Citra Jagung

3.2.2 Mengubah Citra RGB Menjadi Citra Keabuan

Merubah citra menjadi keabuan dilakukan untuk dapat melanjutkan ke tahap selanjutnya yaitu Segmentasi citra. Apabila citra bukan keabuan, maka tidak bisa dilakukan tahap selanjutnya. Berikut adalah hasil dari mengubah citra jagung menjadi citra keabuan.



Gambar 4. Hasil Mengubah Citra Keabuan

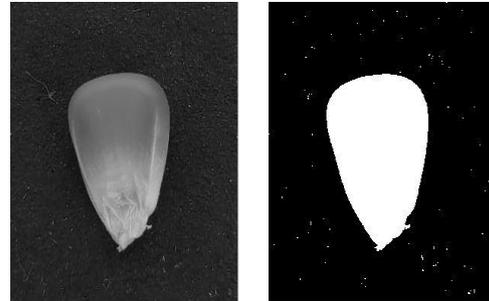
3.3 Segmentasi Citra

Segmentasi citra terdiri dari dua tahap yaitu thresholding dengan opening dan closing morfologi. Tahapan lebih lanjut sebagai berikut:

3.3.1 Thresholding

Metode ini merupakan proses pemisahan antara objek dengan background. Proses thresholding pada matlab menggunakan fungsi “*imbinarize*” diikuti dengan nilai ambang 0-255 sesuai kebutuhan citra. Setelah menjadi citra keabuan pada tahap sebelumnya, lalu diimplementasikan nilai

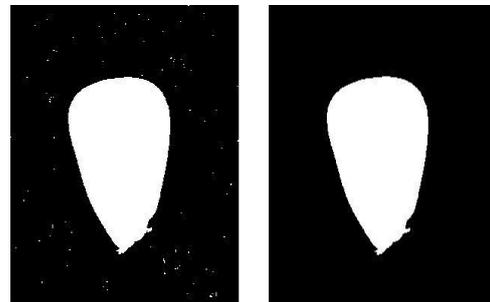
ambang sebesar 80 pada seluruh citra. Berikut adalah hasil dari proses *thresholding*.



Gambar 5. Hasil Proses *Thresholding*

3.3.2 Opening dan Closing Morfologi

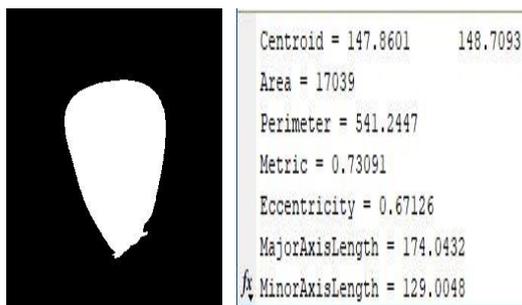
Operasi opening dan closing morfologi berguna untuk menghaluskan kontur serta menghilangkan noise pada citra. Berikut adalah hasil dari *opening* dan *closing morfologi*.



Gambar 5. Hasil *Opening* dan *Closing Morfologi*

3.4 Ekstraksi Ciri

Ekstraksi ciri bentuk akan menghasilkan beberapa nilai. Nilai yang dijadikan parameter untuk proses klasifikasi yaitu nilai *Area*, *Parimeter eccentricity*, *metric*, *width* dan *Length*. Berikut adalah hasil dari ekstraksi ciri yang sudah dilakukan proses *opening* dan *closing morfologi*.



Gambar 6. Hasil Ekstrasi Ciri

Gambar 6. ialah keluaran nilai yang dihasilkan dari citra gambar yang sudah dilakukan proses *opening* dan *closing morfologi*. Untuk membuktikan nilai *area*, *parameter*, *eccentricity*, *metric*, *width* dan *length* yang akan dijadikan parameter, dapat dilakukan dengan cara hitung manual. Adapun tabel 1. dibawah ini merupakan contoh dari hasil nilai ekstraksi ciri pada setiap citra jagung.

Tabel 2. Nilai Hasil Ekstrasi Ciri

N	Are	Perim	Metr	Eccen	Widt	Leng
o	a	eter	ic	ricity	h	th
1	17039	541.2447	0.73091	0.67126	174.0432	129.0048
2	15802	534.0732	0.69618	0.76034	179.3151	116.4704
3	17133	510.86	0.82497	0.59479	165.7227	133.221
4	15852	522.2742	0.73029	0.67688	166.8651	122.8291
5	19016	530.2742	0.84982	0.41932	163.5474	148.4748
6	15332	473.3036	0.86006	0.56572	154.0637	127.0406
...
2						
5	12900	442.9188	0.82633	0.64836	148.1718	112.8084
0						

3.5 Klasifikasi K-Nearest Neighbor

Parameter yang digunakan pada proses klasifikasi yaitu *area*, *parameter*, *metric*, *eccentricity*, *length* dan *width* yang bertujuan untuk memudahkan selama proses klasifikasi. Setelah nilai-nilai berhasil di *export* selanjutnya dilakukan

pemberian label terhadap setiap data. Butir jagung yang merupakan varietas jagung mutiara diberi label “Ya”. Sedangkan butir jagung yang merupakan varietas jagung lain diberi label “Tidak”. Proses klasifikasi dilakukan dengan bantuan tools *Microsoft excel* dan *Rstudio*. Sebelum melakukan pembagian data *Training* dan data *Testing*, nilai-nilai yang dihasilkan dari hasil ekstrasi ciri tersebut memiliki rentang nilai yang terlalu jauh. Hal tersebut dapat mempengaruhi hasil evaluasi dari proses klasifikasi yang dilakukan. Oleh karena itu perlu dilakukan proses transformasi data dengan melakukan normalisasi data. Normalisasi data dilakukan dengan merubah nilai-nilai tersebut sehingga memiliki rentang nilai antara 0 sampai 1 dengan menggunakan rumus sebagai berikut:

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (8)$$

Keterangan:

- y = nilai hasil normalisasi
- x = nilai pada data
- x_{min} = nilai terkecil pada data
- x_{max} = nilai terbesar pada data

Tabel 3. Nilai Max dan Min pada Data

Keter	Ar	Peri	Met	Eccen	Widt	Len
angan	ea	mete	ric	tricity	th	gh
Nilai						
Maks	2229	5969016	0.4890	0.78588	196.4221	154.1983
imum	5		1		1	3
Nilai						
Mini	7208	3266346	0.8882	0.15527	75.1614	80.3168

mum

Berikut merupakan salah satu contoh normalisasi data parameter M1 menggunakan rumus tersebut yaitu:

$$y = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$$y = \frac{541.2447 - 326.6346}{596.9016 - 326.6346}$$

$$y = 0.794066978$$

Berikut ini merupakan hasil normalisasi terhadap data nilai ekstrasi ciri:

Tabel 4. Hasil Normalisasi Ekstrasi Ciri

No	nama File	Area	Parameter	Metric	Eccentricity	Width	Length	Ket
1	M1	0,6 516 206 01	0,7 940 669 78	0,6 059 771 04	0,8 182 394 82	0,8 154 480 39	0,6 590 012 38	Ya
2	M2	0,5 696 294 82	0,7 675 321 07	0,5 189 759 26	0,9 594 995 32	0,8 589 237 9	0,4 893 457 77	Ya
3	M3	0,6 578 511 3	0,6 816 422 28	0,8 416 042 49	0,6 969 759 44	0,7 468 314 14	0,7 160 682 99	Ya
...
250	K30	0,1 900 311 53	0,2 158 084 41	0,9 151 281 34	0,3 387 672 25	0,3 552 008 19	0,4 003 356 73	Tidak

data *Training* dan data *Testing* pada penelitian ini menggunakan rasio pembagian data 7:3. Pada rasio pembagian data 7:3 dari total 250 record data, dimana 70% merupakan data *Training* dan 30% merupakan data *Testing*. Jumlah record pada data *Training* sebanyak 176 sedangkan pada data *Testing* sebanyak 74 record data. Berikut ini merupakan contoh dari pemisahan data *Training* dan data *Testing* dengan rasio perbandingan 7:3 menggunakan *Rstudio*:

Tabel 5. Data *Training*

No	nama File	Area	Parameter	Metric	Eccentricity	Width	Length	Ket
1	M1	0,6 516 206 01	0,7 940 669 78	0,6 059 771 04	0,8 182 394 82	0,8 154 480 39	0,6 590 012 38	Ya
2	M2	0,5 696 294 82	0,7 675 321 07	0,5 189 759 26	0,9 594 995 32	0,8 589 237 9	0,4 893 457 77	Ya
3	M3	0,6 578 511 3	0,6 816 422 28	0,8 416 042 49	0,6 969 759 44	0,7 468 314 14	0,7 160 682 99	Ya
...
176	K30	0,1 900 311 53	0,2 158 084 41	0,9 151 281 34	0,3 387 672 25	0,3 552 008 19	0,4 003 356 73	Tidak

Setelah melakukan normalisasi data pada data nilai ekstrasi ciri, maka proses selanjutnya melakukan pembagian data yang terdapat dua file hasil ekstrasi ciri yaitu file yang terdiri dari data *Training* dan file yang terdiri dari data *Testing*. Jumlah

Tabel 6. Data Testing

No	Nama File	Ar ea	Pa ri me ter	M etric	Ec ce ntr icity	Wi dt h	Le ng th	Ke t
1	M4	0.57294	0.72387	0.60442	0.82715	0.75625	0.57541	Ya
2	M5	0.78266	0.75347	0.90385	0.41872	0.78953	0.92213	Ya
3	M6	0.53847	0.54268	0.92950	0.65087	0.65068	0.63154	Ya
...
74	K23	0.17789	0.23299	0.80477	0.77911	0.48162	0.25368	Tidak

Setelah didapat data *training* dan data *testing* langkah selanjutnya yaitu, perhitungan algoritma *K-Nearest Neighbor* dengan menggunakan rumus *euclidian distance*. Berikut ini langkah-langkah dalam perhitungan algoritma *K-Nearest Neighbor*:

1. Menentukan nilai K. K atau jumlah tetangga terdekat yang digunakan penelitian ini adalah K=3, k=5, dan K=7.
2. Menghitung jarak *Euclidean Distance* dengan nilai diambil secara acak, kemudian memilih data M4 pada data *testing*. Berikut tabel perhitungan *euclidian distance*.

Tabel 7. Perhitungan Euclidian Distance pada Data M4

Nama	Perhitungan	Hasil	Rank
M1 - M4	$\sqrt{\begin{matrix} (0,6516206 \\ (0,6059771 \\ (0,815448 \end{matrix}}$	1,1080657	5
M2 - M4	$\sqrt{\begin{matrix} (0,5696294 \\ (0,518975 \\ (0,85892 \end{matrix}}$	1,1383939	16
M3 - M4	$\sqrt{\begin{matrix} (0,657851 \\ (0,8416042 \\ (0,746831 \end{matrix}}$	1,072961	47
....
K30 - M4	$\sqrt{\begin{matrix} (0,1900311 \\ (0,915128 \\ (0,35520 \end{matrix}}$	0,9650411	147

3. Mengurutkan objek-objek tersebut berdasarkan hasil nilai *euclidian distance* terkecil hingga terbesar. Perhatikan tabel berikut ini.

Tabel 8. Pengurutan Jarak pada Data M4

Nama	Euclidean Distance	Rank	Keterangan
M48	0,088028	1	Ya
M122	0,1231649	2	Ya
M81	0,1301173	3	Ya
M71	0,1357565	4	Ya
M1	0,1472755	5	Ya
M105	0,1528787	6	Ya
M75	0,1536196	7	Ya
...
KP15	1,3077864	176	Tidak

4. Tentukan nilai K untuk mengetahui jumlah tetangga pada objek. Jumlah K yang digunakan dalam penelitian ini yakni 3, 5, 7. Berikut adalah contoh dari jumlah tetangga terdekat terhadap M4. Perhatikan tabel dibawah ini.

Tabel 9. Penentuan Tetangga Terdekat pada Data M4

Nama	Eclidean Distance	Rank	Keterangan	Tetangga
M48	0,088028	1	M48	Ya
M122	0,1231649	2	M122	Ya
M81	0,130173	3	M81	Ya
M71	0,1357565	4	M71	Tidak
M1	0,1472755	5	M1	Tidak
M105	0,1528787	6	M105	Tidak
M75	0,1536196	7	M75	Tidak
...
KP15	1,3077864	176	KP15	Tidak

5. Hasil daripada klasifikasi diperoleh dari jumlah banyaknya K tetangga yang terdekat. Hasil keseluruhan klasifikasi terhadap data testing dapat dilihat pada tabel di bawah ini.

Tabel 10. Hasil Klasifikasi K = 3, 5 dan 7

No	Nama File	Keterangan	K = 3	K = 5	K = 7
1	M4	Ya	Ya	Ya	Ya
2	M5	Ya	Ya	Ya	Ya
3	M6	Ya	Ya	Ya	Ya
...
75	K23	Tidak	Tidak	Tidak	Tidak

3.6 Evaluasi

Setelah melakukan fase build model, selanjutnya dari setiap build model yang telah dilakukan akan diketahui hasilnya. Dari hasil yang didapat nantinya akan ditentukan model mana yang memiliki hasil terbaik. Berikut ini adalah hasil dari model skenario K=3 dengan pembagian data menggunakan *percentage split* dengan *Stratified random sampling* pada rasio pembagian data latih dan data uji 7:3. Perhatikan tabel *confusion matrix* dibawah ini.

Tabel 11. Confusion Matrix 7:3 dengan K=3

Predictiton	Actual		
	Tidak	Tidak 34 (TP)	Ya 2 (FP)
Ya	3 (FN)	35 (TN)	

Setelah diketahui hasil dari tabel *Confusion Matrix*, maka selanjutnya ialah mencari nilai dari *accuracy*, *recall*, *specificity*, *precision*, *error rate* dan *f-measure* dengan cara yang sudah dijelaskan pada penjelasan sebelumnya. Berikut dibawah ini merupakan hasil perhitungan *accuracy*, *recall*, *specificity*, *precision*, *error rate* dan *f-measure* pada *confusion matrix* 7:3 dengan K=3.

$$Accuracy: \frac{(TP+TN)}{(TP+FP+FN+TN)} \times 100\% =$$

$$\frac{(34 + 35)}{(34 + 2 + 3 + 35)} \times 100\% = 93.24\%$$

$$Recall: \frac{(TP)}{(TP+FN)} \times 100\% =$$

$$\frac{(34)}{(34 + 3)} \times 100\% = 91.89\%$$

$$Precision: \frac{(TP)}{(TP+FP)} \times 100\% =$$

$$\frac{(34)}{(34 + 2)} \times 100\% = 94.44\%$$

$$Specificity: \frac{(TN)}{(TN+FP)} \times 100\% =$$

$$\frac{35}{35 + 2} \times 100\% = 94.59\%$$

$$Error Rate: \frac{(FP+FN)}{(TP+FP+FN+TN)} \times 100\% =$$

$$\frac{(2 + 3)}{(34 + 2 + 3 + 35)} \times 100\% = 6.76\%$$

$$F\text{-measure: } 2x \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} = \frac{2 \times \frac{0.9444 \times 0.9189}{0.9444 + 0.9189}}{2} = 0.9315$$

3.6.1 Hasil Evaluasi

Berdasarkan nilai yang didapatkan dari *confussion matrix*, maka dilakukan perhitungan untuk mendapatkan nilai *accuracy*, *recall*, *specificity*, *precision*, *error rate* dan *f-measure* yang dapat dilihat pada tabel di bawah ini.

Tabel 12. Hasil *accuracy*, *recall*, *specificity*, *precision*, *error rate* dan *f-measure*

	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>	<i>Specificity</i>	<i>Error Rate</i>	<i>F-measure</i>
K3	93.24 %	91.8 %	94.44 %	94.59 %	6.76 %	0.9315
K5	89.19 %	89.1 %	89.19 %	97.30 %	10.8 %	0.89
K7	87.84 %	78.38 %	96.67 %	85.37 %	12.16 %	0.86

Pada **Tabel 12.** mepresentasikan bahwa nilai akurasi tertinggi terdapat pada hasil klasifikasi menggunakan Algoritma *K-Nearest Neighbor* dengan nilai $k=3$ yaitu sebesar 93.24%. Begitupun dengan nilai *recall* tertinggi terdapat pada hasil klasifikasi dengan nilai $k=3$ yaitu sebesar 91.89%. Sedangkan, nilai presisi tertinggi terdapat pada hasil klasifikasi dengan nilai $k=7$ yaitu sebesar 96.67%. Nilai *specificity* tertinggi terdapat pada hasil klasifikasi dengan nilai $k=5$ yaitu sebesar 97.30%. Sedangkan nilai *error rate* tertinggi terdapat pada hasil klasifikasi dengan nilai $k=7$ yaitu sebesar 12.16% dan nilai *f-measure* tertinggi terdapat pada hasil klasifikasi dengan nilai $k=3$ yaitu sebesar 0.9315. Hasil evaluasi pada nilai $k = 3$ sampai $k = 7$ mengalami penurunan pada nilai akurasi sehingga proses klasifikasi dihentikan pada nilai $k = 7$.

3.7 Pembahasan

Berdasarkan hasil dari penelitian identifikasi varietas jagung Mutiara menggunakan Algoritma *K-Nearest Neighbor* dengan menggunakan nilai $k = 3, 5, 7$ dan perhitungan jarak menggunakan rumus *euclidean distance*. Nilai yang digunakan untuk perhitungan jarak adalah hasil dari ekstraksi ciri bentuk setiap butir citra jagung mutiara dan butir citra varietas jagung lain. Nilai ekstraksi ciri yang digunakan yaitu nilai *area*, *parimeter*, *metric*, *eccentricity*, *length* dan *width*. Dan dari hasil evaluasi tersebut nilai k optimal yang digunakan untuk identifikasi varietas jagung mutiara menggunakan algoritma *K-Nearest Neighbor* adalah $k = 3$. Hal ini dikarenakan nilai *accuracy*, *recall* dan *f-measure* tertinggi dibandingkan model yang lain, serta memiliki nilai *error rate* terendah.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat diambil kesimpulan sebagai berikut:

1. Untuk membedakan varietas jagung mutiara dengan varietas jagung lain dapat dilakukan dengan pengolahan citra digital. Salah satu hal yang membedakan varietas jagung mutiara dengan varietas jagung lainnya adalah dari segi bentuk. Dengan pengolahan citra digital ciri-ciri bentuk suatu citra jagung mutiara dapat diekstraksi menjadi nilai-nilai khusus yang mencirikan suatu citra. Nilai-nilai khusus dari hasil ekstraksi ciri bentuk jagung mutiara berdasarkan data citra digital diantaranya *area*, *parimeter*, *length*, *width*, *metric* dan *eccentricity*.
2. Identifikasi varietas jagung mutiara dilakukan dengan menggunakan metode klasifikasi Algoritma *K-Nearest Neighbor*. Dan nilai *accuracy* tertinggi terdapat pada $k = 3$ yaitu sebesar 93.24% dengan nilai *F-measure* 0.9315 dan *error rate* sebesar 6.76% yang

menjadikan ini adalah model terbaik dalam melakukan identifikasi jagung mutiara.

3. Evaluasi keberhasilan Algoritma *K-Nearest Neighbor* dalam mengidentifikasi varietas jagung mutiara dapat diketahui berdasarkan nilai akurasi, presisi, recall, specificity, error rate dan f-measure. Secara keseluruhan hasil evaluasi dengan nilai $k = 3, 5, 7$ memiliki nilai akurasi paling rendah sebesar 87.84% dan akurasi tertinggi sebesar 93.24%. Sedangkan nilai recall paling rendah sebesar 78.38% dan tertinggi sebesar 91.89%. Nilai presisi terendah sebesar 89.19% dan tertinggi sebesar 96.67%. Specificity terendah sebesar 85.37% dan tertinggi sebesar 97.30%. Error rate terendah sebesar 6.76% dan tertinggi sebesar 12.16%. Yang terakhir adalah f-measure terendah sebesar 0.8657 dan tertinggi sebesar 0.9315. Dan dari hasil evaluasi tersebut nilai k optimal yang digunakan untuk identifikasi varietas jagung mutiara menggunakan algoritma *K-Nearest Neighbor* adalah $k = 3$.

5. SARAN

Dari penelitian yang telah dilakukan, saran yang dapat diberikan untuk penelitian selanjutnya adalah:

1. Data citra yang digunakan bisa menggunakan data citra yang lebih banyak.
2. Identifikasi varietas jagung mutiara dengan menggunakan metode klasifikasi dapat dicoba dengan menggunakan algoritma lain selain *K-Nearest Neighbor*.
3. Rasio pembagian data dapat menggunakan perbandingan selain 7:3.

4. Pengujian hasil dari klasifikasi dapat menggunakan metode selain *Confusion Matrix*.
5. Menerapkan metode identifikasi varietas jagung mutiara menggunakan Algoritma *K-Nearest Neighbor* kedalam program.

DAFTAR PUSTAKA

- [1] K. Kementerian Pertanian Republik Indonesia, "Rencana Strategis Kementerian Pertanian 2020-2024," *Rencana Strategis Kementerian Pertanian 2020-2024*, vol. 7, no. 2. pp. 1–16, 2020.
- [2] Kementerian Pertanian, *Rencana Strategis Kementerian Pertanian 2015-2019 : Edisi Revisi*. 2016.
- [3] Subairi, Rahmadwati, and E. Yudaningtyas, "Implementasi Metode k-Nearest Neighbor pada Pengenalan Pola Tekstur Citra Saliva untuk Deteksi Ovulasi," *J. EECCIS*, vol. 12, no. 1, pp. 9–14, 2018.
- [4] S. R. Sulistiyanti, F. A. Setyawan, and M. Komarudin, *Pengolahan Citra Dasar dan Contoh Penerapannya*. Yogyakarta: TEKNOSAIN, 2016.
- [5] A. A. I. Wiratmaka, I. F. Rozi, and R. A. Asmara, "Klasifikasi Kualitas Tanaman Cabai Menggunakan Metode Fuzzy K-Nearest Neighbor (Fknn)," *J. Inform. Polinema*, vol. 3, no. 3, pp. 1–5, 2017.
- [6] Y. F. Rachman, Kusri, and H. Al Fatta, "Klasifikasi Citra Digitalretina Penderita Diabetes Retinopati Menggunakan Metode Euclidean," *J. Comput. Inf. Technol.*, vol. 3, no. 2, pp. 75–82, 2020.
- [7] Rusli and M. Nasir, "Klasifikasi Jenis Mangga Berdasarkan Bentuk Daun Menggunakan Metode K-Nearest Neighbor," *J. Infomedia*, vol. 3, no. 2, pp. 87–91, 2018.
- [8] A. S. Hassoon and L. F. Jalil, "Classification of Iraqi Anber Rice by Using Image Processing and

- KNN Algorithm,” *Iraqi J. Sci.*, vol. 58, no. 2A, pp. 716–725, 2017.
- [9] M. Orisa and T. Hidayat, “Analisis Teknik Segmentasi Pada Pengolahan Citra,” *J. Mnemon.*, vol. 2, no. 2, pp. 9–13, 2019.
- [10] A. R. Ginanjar, “Penggabungan Ciri Tekstur dan Geometri Citra Untuk Identifikasi Tumbuhan Obat Berbasis Mobile,” 2015.
- [11] P. N. Andono, T. Sutojo, and Muljono, *Pengolahan Citra Digital*. Yogyakarta: Penerbit ANDI, 2017.
- [12] R. Siringoringo, “Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor,” *J. ISD*, vol. 3, no. 1, pp. 44–49, 2018.
- [13] S. D. Nugraha, R. R. M. Putri, and R. C. Wihandika, “Penerapan Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentukan Status Gizi Balita,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 9, pp. 925–932, 2017.
- [14] D. T. Larose, *DISCOVERING KNOWLEDGE IN DATA: An Introduction to Data Mining*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2005.
- [15] J. Han, M. Kamber, and J. Pei, *DATA MINING: Concepts and Techniques*, Third Edit. USA: Elsevier Inc., 2012.
- [16] M. Hossin and M. N. Sulaiman, “A Review on Evaluation Metrics for Data Classification Evaluations,” *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 1–11, 2015.