



Penerapan Data Mining Dengan Metode K-Nearest Neighbor Terhadap Klasifikasi Sarang Walet

Muhammad Ismail, Renaldi Yulvianda*

Fakultas Ilmu Komputer, Teknik Informatika, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: ¹maelkhanz@gmail.com, ²*Ryulvianda@gmail.com

Email Penulis Korespondensi: Ryulvianda@gmail.com

Abstrak—Desa Sungai Benuh merupakan salah satu daerah yang banyak membuat rumah walet karena mampu menghasilkan sarang walet dengan jumlah yang banyak sehingga tujuan penelitian ini dilakukan adalah untuk membuat klasifikasi dengan penerapan data mining untuk melihat tingkat kualitas sarang walet, sehingga nantinya menjadi referensi dalam membantu para pembeli dan penjual memperoleh hasil yang sesuai dan menjaga daya jual sarang walet tersebut. Hal ini juga didasari oleh masalah yang sering dihadapi yaitu penjual dan pembeli tidak memiliki standar penilaian yang tetap ketika transaksi berlangsung sehingga muncul penilaian sepihak. Selain itu, terdapat perbedaan kualitas dan kuantitas dimusim yang berbeda. Saat musim hujan sarang walet yang diperoleh lebih besar, putih, bersih dan banyak sedangkan saat kemarau diperoleh hasil sebaliknya. Hasil klasifikasi menggunakan metode k-nearest neighbor dengan software weka menunjukkan akurasi 90% terhadap 45 dari 50 sampel data, termasuk sampel data pembeding atau sampel data baru dengan menggunakan nilai $k = 7$ dengan atribut kebersihan, warna, ukuran, bentuk dan masa panen yang dikategorikan “Bagus” atau “Buruk”. Evaluasi hasil dengan confusion matrix diperoleh hasil akurasi yaitu 80%, precision 80.49, recall 94.29% dan F1 score 86.84%. Jadi, penelitian ini berhasil dilakukan dengan hasil klasifikasi yang tinggi sehingga dapat menjadi referensi guna membantu para pembeli dan penjual memperoleh kesepakatan bersama saat transaksi dan menjaga daya jual sarang walet tersebut.

Kata Kunci: Penerapan Data Mining; K-Nearest Neighbor; Sarang Walet; Confusion Matrix; Sungai Benuh

Abstract—Sungai Benuh Village is one of the areas where many swallow houses are made because it is able to produce a large number of swallow nests so that the purpose of this study was to make a classification by applying data mining to see the quality level of swallow nests, so that later it will become a reference in helping buyers and the seller obtains appropriate results and maintains the selling power of the swallow's nest. This is also based on a problem that is often encountered, namely sellers and buyers do not have a fixed standard of evaluation when a transaction takes place, so that unilateral judgments appear. In addition, there are differences in quality and quantity in different seasons. During the rainy season, swallow nests are larger, white, clean and numerous, while during the dry season, the opposite results are obtained. Classification results using the k-nearest neighbor method with Weka software show 90% accuracy for 45 out of 50 data samples, including comparative data samples or new data samples using a value of $k = 7$ with categorized attributes of cleanliness, color, size, shape and harvest time “Good” or “Bad”. Evaluation of the results with the confusion matrix results obtained accuracy of 80%, precision 80.49, recall 94.29% and F1 score 86.84%. So, this research was successfully carried out with high classification results so that it can be a reference to help buyers and sellers obtain a mutual agreement during transactions and maintain the selling power of the swallow's nest.

Keywords: Application of Data Mining; K-Nearest Neighbor; Swallow's Nest; Confusion Matrix; Sungai Benuh

1. PENDAHULUAN

Peternakan sebagai salah satu usaha yang dilakukan baik untuk memelihara ataupun budidaya dengan harapan memperoleh keuntungan sesuai yang ditargetkan dengan mengupayakan hasil produksinya menjadi optimal [1]. Indonesia menjadi tempat budidaya dan pengekspor sarang walet terbesar kenegara lain yaitu 75% dikarenakan kondisinya yang sesuai untuk perkembangbiakan burung walet [2].

Burung walet atau *Collocalia vestita* merupakan jenis burung dengan sayap meruncing, bulu berwarna hitam, dan ekor Panjang yang menyukai tempat gelap dan lembab seperti gua ataupun bangunan gedung serta sebagai burung pemakan serangga [3]. Spesies ini hidup berkelompok dengan membuat sarang dari air liurnya dan sarang tersebut memiliki khasiat di bidang kesehatan dan harga jual yang sangat tinggi [4]. Sarang walet memiliki Teknik panen yaitu dengan cara panen rampas (masih ada walet yang bersarang) dan panen menetap (walet membuat sarang baru). Namun perlu dilakukan dengan benar agar tidak mengganggu dan membuat walet tersebut pindah ke Gedung lain.

Desa sungai benuh merupakan salah satu daerah yang banyak membuat rumah walet atau yang disebut dengan Gedung walet dan mampu menghasilkan sarang walet dengan jumlah yang banyak. Selain itu, ketertarikan investor pada pembangunan gedung untuk budidaya burung walet semakin tinggi seiring dengan besarnya potensi ekonomi yang ada. Pembeli dari luar daerah lain berdatangan untuk membeli hasil panen sarang walet tersebut, namun terdapat masalah yang sering muncul yaitu penjual dan pembeli tidak memiliki standar penilaian yang tetap ketika transaksi berlangsung sehingga muncul penilaian sepihak. Misal, pembeli akan membayar mahal jika kondisi sarang walet tersebut bersih, berwarna putih, ukuran besar dan kering. Selain dari yang disebutkan akan dibeli dengan harga murah. Selain itu, terdapat perbedaan kualitas dan kuantitas dimusim yang berbeda. Saat musim hujan sarang walet yang diperoleh lebih besar, putih, bersih dan banyak sedangkan saat musim kemarau atau jarang hujan maka sarang walet yang didapat lebih kecil, hitam, penuh bulu dan sering pecah serta hasil panennya lebih sedikit.



Melalui penelitian ini, peneliti mencoba untuk membuat pengklasifikasian kualitas sarang walet berdasarkan kriteria kebersihan, bentuk, warna, masa panen dan ukuran sehingga dapat menjadi tolak ukur saat transaksi dilakukan dengan pendekatan data mining. Data mining adalah proses dalam menggali informasi yang belum diketahui dari kumpulan data yang besar dan mampu melakukan ekstraksi data tersebut sehingga menghasilkan pengetahuan yang memiliki nilai tambah [5] [6], [7]. Salah satu teknik yang dapat digunakan dalam data mining adalah teknik klasifikasi yang merupakan proses yang digunakan untuk menemukan kumpulan pola untuk memisahkan kelas data berdasarkan hubungan data dengan sampel data [8] [9]. Klasifikasi memberikan kelompok kepada setiap keadaan yang berisi class atribut [9]

Beberapa peneliti sebelumnya sudah banyak yang melakukan klasifikasi terhadap suatu objek dengan menggunakan metode K-Nearest Neighbor. Seperti halnya dengan peneliti [10] melakukan klasifikasi produksi ayam broiler untuk melihat tingkat keberhasilan produksi ayam broiler tersebut. Hasil evaluasi yang diperoleh setelah mengukur hasil klasifikasi dengan k-nearest neighbor menggunakan confusion matrix yaitu tingkat akurasi sebesar 86.49% pada nilai $k = 3$, presisi 75% dan recall 70.21%.

Peneliti [11] berhasil melakukan klasifikasi telur puyuh dengan sistem yang ada dengan tingkat akurasi sebesar 77.78% berdasarkan kriteria baik, sedang dan buruk. Penelitian [12] mengklasifikasikan hasil panen jamur tiram dengan K-Nearest Neighbor diperoleh hasil Mean absolute error sebesar 0.1419, Root mean squared error sebesar 0.2111, Relative absolute error sebesar 36.2177% dan Root relative squared error sebesar 48.002%. Penelitian [13] berhasil melakukan klasifikasi terhadap biji kopi yang rusak dengan membuat sistem dengan hasil keakuratan 90%.

Penelitian lain yaitu [14] melakukan percobaan untuk mengetahui pengaruh penerapan fitur seleksi Principal Component Analysis (PCA) terhadap performa algoritma K-Nearest Neighbor (KNN) guna memprediksi hasil produksi agrikultur menghasilkan aplikasi dengan akurasi paling tinggi sebesar 99,64%.

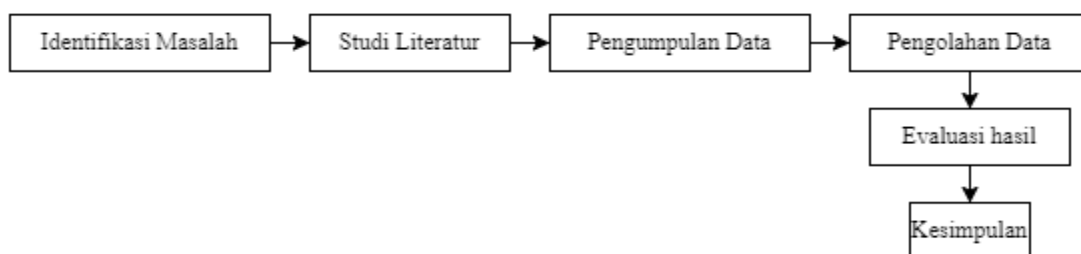
Merujuk pada penelitian yang sudah dilakukan sebelumnya, belum ada penelitian yang secara khusus membahas hasil panen pada sarang walet dan hanya terfokus pada hasil peternakan atau budidaya lainnya. Selain itu, penelitian ini dilakukan dengan menggunakan data set yang berbeda yaitu dari segi jumlah data dan atribut sarang walet serta melakukan pengujian yang lebih kompleks untuk memvalidasi hasil dari klasifikasi (K-Nearest Neighbor) yang telah diperoleh dengan confusion matrix mulai dari accuracy, precision, recall, specificity atau true negative rate dan f-score.

Dari uraian permasalahan diatas, penelitian ini bertujuan untuk melakukan proses klasifikasi terhadap kualitas sarang walet dengan metode K-Nearest Neighbor dalam data mining. Hasil penelitian ini diharapkan mampu menjadi referensi dalam membantu para pembeli dan penjual sehingga memperoleh hasil yang sesuai kesepakatan bersama dan menjaga daya jual sarang walet tersebut.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan untuk mengetahui klasifikasi dari hasil panen budidaya sarang burung walet yang ada di Desa Sungai Benuh dengan menggunakan metode K-Nearest Neighbor sehingga dapat diketahui kualitas terbaik yang telah diperoleh selama ini. Berikut tahapan penelitian yang dibuat agar menjadikan penelitian lebih terstruktur:



Gambar 1. Tahapan Penelitian

a. Identifikasi Masalah

Identifikasi masalah dilakukan dengan cara observasi ke beberapa tempat pemilik sarang burung walet dan diwawancara terkait permasalahan yang sering dialami sehingga dapat dijabarkan bahwa penjual dan pembeli tidak memiliki standar penilaian yang tetap ketika transaksi berlangsung sehingga muncul penilaian sepihak. Misal, pembeli akan membayar mahal jika kondisi sarang walet tersebut bersih, berwarna putih, ukuran besar dan kering. Selain dari yang disebutkan akan dibeli dengan harga murah. Selain itu, terdapat perbedaan kualitas dan kuantitas dimusim yang berbeda. Saat musim hujan sarang walet yang diperoleh lebih besar, putih, bersih dan banyak sedangkan saat kemarau sarang walet yang didapat lebih kecil, hitam, penuh bulu dan sering pecah serta hasil panennya lebih sedikit.

**b. Studi Literatur**

Teori pendukung sebagai pelengkap dalam penelitian ini menggunakan jurnal dan buku yang diambil dari google scholar dan e-book yang terkait dengan klasifikasi, data mining, sarang walet, metode k-nearest neighbor dan lain sebagainya

c. Pengumpulan Data

Penelitian ini menggunakan data yang dikumpulkan melalui observasi dan wawancara kepada beberapa pemilik sarang burung walet yang ada di Desa Sungai Benuh. Data yang diperoleh dijadikan sampel pengujian dengan jumlah 49 sampel sarang burung walet dan 1 data sampel baru atau sampel pembandingan. Data tersebut akan diklasifikasikan dengan beberapa atribut yaitu berdasarkan dari kebersihan, bentuk, warna, masa panen dan ukuran.

d. Pengolahan Data

Penelitian ini menggunakan metode k-nearest neighbor dalam data mining untuk mengklasifikasikan sarang walet dengan menggunakan sampel data dan sampel data baru atau pembandingan yang akan diolah dengan software weka.

e. Evaluasi Hasil

Evaluasi hasil ini dilakukan untuk melihat tingkat hasil pengukuran dari metode k-nearest neighbor baik nilai akurasi, presisi maupun nilai recallnya dengan menggunakan confusion matrix yang terdapat di software rapidminer untuk melihat tingkat keberhasilan dari hasil klasifikasi k-nearest neighbor.

2.2 Metode K-Nearest Neighbor

Metode k-nearest neighbor adalah metode yang menerapkan algoritma supervised learning untuk klasifikasi suatu objek berdasarkan data jarak tetangga terdekatnya [15]. Algoritma k-nearest neighbor terbagi menjadi dua jenis, yaitu pembelajaran terkontrol dan tidak terkontrol. Pembelajaran terkontrol Algoritma bertujuan untuk memperoleh pola baru, sedangkan algoritma pembelajaran yang tidak terkontrol bertujuan untuk mempertahankan pola data [16]. Berikut langkah-langkah yang ada pada metode k-nearest neighbor [17]:

a. Menentukan hasil Menentukan nilai k (jumlah tetangga yang paling dekat) yang akan digunakan.

b. Menghitung nilai kedekatan data berdasarkan perhitungan Euclidean terhadap data latih dengan persamaan:

$$D(x - y) = ||x - y||_2 = \sqrt{\sum_{j=1}^N |x^j - y^j|^2} \quad (1)$$

c. Mengurutkan hasil jarak dari urutan nilai tertinggi ke nilai terendah

d. Menghitung jumlah setiap kelas berdasarkan k tetangga terdekat

e. Kelas mayoritas dijadikan sebagai kelas baru lagi bagi data uji.

2.3 Sarang Walet

Sarang walet yang bersih, berwarna putih kapas, kondisi utuh dan bentuknya mangkok adalah sarang walet yang berkualitas dengan kriteria super [18]. Harga jual sarang walet sangat mahal meskipun terbuat dari air liur burung walet dengan jenis burung Collocalia fuchiphaga (sarang putih), collocalia esculenta (dibudidayakan sebagai burung seriti) dan Collocalia maxima (sarang hitam) [19]. Selain itu, sarang walet dapat dijadikan sebagai obat karena memiliki khasiat yang baik yaitu dikonsumsi untuk menjaga daya tahan tubuh, sesak napas dan asma [20].

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini hasil dan pembahasan berisikan tentang penggunaan metode K-Nearest Neighbour untuk melihat klasifikasi yang diolah di software WEKA 3.8.6 dan hasilnya akan dievaluasi dengan menggunakan confusion matrix yang diolah di software rapidminer untuk melihat tingkat akurasi, presisi dan recall yang diperoleh dari hasil k-nearest neighbor.

3.1 Pengolahan Data**3.1.1 Sampel Data Sarang Walet**

Penelitian ini menggunakan data sarang walet yang didapat dari salah satu pemilik sarang walet di Sungai Benuh yang telah dipilih berdasarkan kriteria tertentu untuk dijadikan sampel yang diuji dengan jumlah data sebanyak 49 sarang walet dan mewakili sarang walet lainnya.

Tabel 1. Sampel Data Sarang Walet

ID Sarang Walet	Warna	Kebersihan	Ukuran (CM)	Bentuk	Masa Panen
SW-01	0	1	5	1	20
SW-02	0	1	6	1	22
SW-03	1	1	6	1	22
SW-04	0	1	7	0	25



ID Sarang Walet	Warna	Kebersihan	Ukuran (CM)	Bentuk	Masa Panen
SW-05	0	1	7	0	25
SW-06	1	1	7	1	25
SW-07	0	1	7	1	25
SW-08	0	1	7	1	25
SW-09	1	1	7.5	0	30
SW-10	1	1	8	1	35
SW-11	0	1	8	1	35
SW-12	0	1	8	1	35
SW-13	0	1	8	1	35
SW-14	1	1	8	0	35
SW-15	1	1	8	0	35
SW-16	1	1	8	1	35
SW-17	1	1	8	1	35
SW-18	1	1	8	1	35
SW-19	1	1	8	1	35
SW-20	0	1	8	1	35
SW-21	1	1	8	1	35
SW-22	1	1	8	1	35
SW-23	0	1	8	1	35
SW-24	0	1	8.5	0	38
SW-25	0	1	9	1	40
SW-26	1	1	9	1	40
SW-27	1	1	9	1	40
SW-28	1	1	9	0	40
SW-29	1	1	9	1	40
SW-30	0	1	9	1	40
SW-31	1	1	9	0	40
SW-32	1	1	9	1	45
SW-33	0	1	9	1	45
SW-34	0	1	9	0	45
SW-35	1	1	9	0	45
SW-36	1	1	9	1	45
SW-37	0	1	9	1	45
SW-38	0	1	9	1	45
SW-39	1	1	9	1	45
SW-40	0	1	9	1	45
SW-41	0	0	10	1	55
SW-42	1	0	10	1	55
SW-43	1	0	10	1	55
SW-44	1	1	10	0	55
SW-45	0	1	10	1	55

Tabel 2. Atribut Sampel Data

Atribut	Kode	Nilai
Bentuk	Sudut	0
	Mangkok	1
Kebersihan	Banyak Bulu	0
	Sedikit Bulu	1
Warna	Putih Kapas	0
	Putih Kecoklatan	1
Ukuran		7 cm
Masa Panen		30 Hari

Berikut tabel 2. yang berisikan atribut untuk kriteria penentuan klasifikasi hasil sarang walet sebagai sampel data yang digunakan dalam penelitian ini. Atribut yang digunakan yaitu bentuk, kebersihan, warna, ukuran dan masa panen. Untuk atribut bentuk, kebersihan dan warna karena memiliki nilai kode berupa text sehingga diubah menjadi data numerik dengan metode label end coding yaitu 0 dan 1.

3.1.2 Klasifikasi Data Dengan K-Nearest Neighbor

Klasifikasi dilakukan untuk mengelompokkan sarang walet ke kategori “Baik” dan “Buruk” berdasarkan atribut “Kebersihan”, “Bentuk”, “Warna”, “Ukuran” dan “Masa Panen”. Parameter K yang digunakan yaitu K = 7 sebagai



tetangga terdekat dari nilai yang dicari. Sedangkan untuk jarak antara sampel data digunakan data baru sebagai pembandingan yang belum diketahui masuk ke kategori baik atau buruk yaitu pada tabel 3 dibawah:

Tabel 3. Sampel Data Baru (Pembandingan)

ID Sarang Walet	Warna	Kebersihan	Ukuran (cm)	Bentuk	Masa Panen (hari)
SW – 50	1	0	11	0	65

adapun hasil perhitungan menggunakan persamaan Euclidean untuk melihat jarak antara sampel data yang ada sebagai berikut:

Tabel 4. Jarak dan Rangking Sampel Data

ID Sarang Walet	Jarak Euclidian	Ranking
SW-01	45,43127	50
SW-02	43,32436	49
SW-03	43,31282	48
SW-04	40,22437	43
SW-05	40,22437	43
SW-06	40,22437	43
SW-07	40,2368	46
SW-08	40,2368	46
SW-09	35,18878	42
SW-10	30,18278	30
SW-11	30,19934	37
SW-12	30,19934	37
SW-13	30,19934	37
SW-14	30,16621	28
SW-15	30,16621	28
SW-16	30,18278	30
SW-17	30,18278	30
SW-18	30,18278	30
SW-19	30,18278	30
SW-20	30,19934	37
SW-21	30,18278	30
SW-22	30,18278	30
SW-23	30,19934	37
SW-24	27,15235	27
SW-25	25,13961	25
SW-26	25,11971	22
SW-27	25,11971	22
SW-28	25,0998	20
SW-29	25,11971	22
SW-30	25,13961	25
SW-31	25,0998	20
SW-32	20,14944	12
SW-33	20,17424	16
SW-34	20,14944	12
SW-35	20,12461	11
SW-36	20,14944	12
SW-37	20,17424	16
SW-38	20,17424	16
SW-39	20,14944	12
SW-40	20,17424	16
SW-41	10,14889	7
SW-42	10,0995	2
SW-43	10,0995	2
SW-44	10,0995	2
SW-45	10,19804	9
SW-46	10,19804	9
SW-47	10,14889	7
SW-48	10,0995	2



ID Sarang Walet	Jarak Euclidian	Ranking
SW-49	10,0995	2

Hasil perhitungan jarak antara sampel data yang diperoleh akan diurutkan dari nilai tertinggi ke nilai terendah yang bisa dilihat dikolom ranking pada tabel 4 diatas. Nilai tertinggi yang diperoleh yaitu 50 dan terendah yaitu 2. Namun yang masuk kedalam kategori tetangga terdekat dari sampel data SW-50 berdasarkan $k = 7$ dengan melihat nilai atribut kebersihan, warna, ukuran, bentuk dan masa panen sebagai berikut :

Tabel 5. Urutan Jarak Terdekat Ke Data Sampel SW-50

ID Sarang Walet	Warna	Kebersihan	Ukuran (cm)	Bentuk	Masa Panen (hari)	Jarak Euclidian	Ranking	Kategori
SW-41	0	0	10	1	55	10,14889157	7	Buruk
SW-42	1	0	10	1	55	10,09950494	2	Buruk
SW-43	1	0	10	1	55	10,09950494	2	Buruk
SW-44	1	1	10	0	55	10,09950494	2	Buruk
SW-47	1	1	10	1	55	10,14889157	7	Bagus
SW-48	1	1	10	0	55	10,09950494	2	Buruk
SW-49	1	0	10	1	55	10,09950494	2	Buruk

Berdasarkan tabel 5 diatas, dapat diketahui bahwa terdapat 7 sampel data yang memiliki jarak terdekat ke sampel data SW-50 yang menjadi sampel data baru atau pembandingan dalam penelitian ini. Hasil klasifikasi yang diperoleh yaitu data sampel SW-50 masuk kedalam kategori “Buruk” karena lebih banyak data sampel yang menghasilkan kategori “Buruk” berdasarkan tetangga terdekat tersebut. Dari ke 7 sampel data tersebut yang masuk kedalam kategori “Bagus” hanya 1 data sampel SW-47 berdasarkan atribut warna yaitu putih kapas (1), kebersihan yaitu sedikit bulu (1), ukuran 10 cm, bentuk mangkok (1) dan masa panen 55 hari. Sedangkan ke 6 data lainnya masuk kedalam kategori “Buruk” dengan atribut warna yaitu putih kecoklatan (0), kebersihan yaitu banyak bulu (0), ukuran 10 cm, bentuk sudut (0) dan masa panen 55 hari.

```

Classifier output

IB1 instance-based classifier
using 7 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          45           90 %
Incorrectly Classified Instances        5           10 %
Kappa statistic                        0.7826
Mean absolute error                    0.0957
Root mean squared error                0.2324
Relative absolute error                22.492 %
Root relative squared error            50.4412 %
Total Number of Instances              50

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.857    0.000    1.000     0.857    0.923     0.802    0.989    0.995    Buruk
                1.000    0.143    0.750     1.000    0.857    0.802    0.989    0.960    Bagus
Weighted Avg.   0.900    0.043    0.925     0.900    0.903     0.802    0.989    0.984

=== Confusion Matrix ===
 a  b  <-- classified as
 30  5 | a = Buruk
  0 15 | b = Bagus
    
```

Gambar 2. Hasil Klasifikasi K-Nearest Neighbour Dari Software Weka

Berikut hasil klasifikasi sarang walet dengan metode k-nearest neighbor yang dikalkulasikan melalui software weka sehingga dapat diketahui bahwa hasil klasifikasi selama 0.00 detik yang dibuat menunjukkan hasil akurasi 90% terhadap 45 dari 50 sampel data yang ada termasuk sampel data pembandingan atau sampel data baru dengan menggunakan nilai $k = 7$. Dengan hasil yang diperoleh dapat disimpulkan jika penggunaan metode k-nearest neighbor terhadap klasifikasi sarang walet berdasarkan atribut yang digunakan berhasil dilakukan. Namun, untuk memperkuat hasil yang diperoleh maka akan dilakukan evaluasi terhadap hasil tersebut menggunakan confusion matrix dengan software rapidminer.

3.2 Evaluasi Hasil Dari Klasifikasi K-Nearest Neighbor

Evaluasi hasil klasifikasi dari metode k-nearest neighbour pada penelitian ini adalah menggunakan confusion matrix yang merupakan pengukuran analisis untuk melihat tingkat keberhasilan dari hasil klasifikasi yang telah diperoleh berdasarkan accuracy, precision, dan recall [21], [22]. Confusion matrix dapat menyatakan hasil klasifikasi jumlah data yang benar dan jumlah data yang salah berdasarkan tabel pengukurannya. Berikut tabel confusion matrix yang dapat dilihat dibawah ini [17]:



Tabel 6. Tabel Confusion Matrix

Actual Class	Assigned Class	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

True Positive (TP) adalah dataset yang diklasifikasikan positif terdeteksi benar dan True Negative (TN) adalah dataset yang diklasifikasikan negative dan terdeteksi benar. Sedangkan False Positive (FP) merupakan dataset yang diklasifikasikan positif yang terdeteksi salah dan False Negative (FN) merupakan dataset yang diklasifikasikan negative dan terdeteksi salah [23].

Penelitian ini menggunakan confusion matrix yang diolah dengan software rapidminer untuk melihat tingkat keberhasilan klasifikasi yang dilakukan dengan menguji hasil akurasi klasifikasi dengan metode k-nearest neighbor. Berikut hasil evaluasi klasifikasi dengan metode k-nearest neighbor menggunakan confusion matrix melalui software rapidminer studio :

Tabel 7. Hasil Accuracy Dari Evaluasi Dengan Confusion Matrix

Accuracy: 80.00 %

	True Buruk	True Bagus	Class Precision
Pred. Buruk	33	8	80.49%
Pred. Bagus	2	7	77.78%
Class recall	94.29%	46.67%	

Tabel 8. Hasil True Positive Confusion Matrix

True_positive: 33.000 (positive class: Buruk)

	True Bagus	True Buruk	Class Precision
Pred. Bagus	7	2	77.78%
Pred. Buruk	8	33	80.49%
Class recall	46.67%	94.29%	

Tabel 9. Hasil True Negative Evaluasi Confusion Matrix

True_negative: 7.000 (positive class: Buruk)

	True Bagus	True Buruk	Class Precision
Pred. Bagus	7	2	77.78%
Pred. Buruk	8	33	80.49%
Class recall	46.67%	94.29%	

Tabel 10. Hasil False Positive Evaluasi Confusion Matrix

false_positive: 8.000 (positive class: Buruk)

	True Bagus	True Buruk	Class Precision
Pred. Bagus	7	2	77.78%
Pred. Buruk	8	33	80.49%
Class recall	46.67%	94.29%	

Tabel 11. Hasil False Negative Evaluasi Confusion Matrix

false_negative: 2.000 (positive class: Buruk)

	True Bagus	True Buruk	Class Precision
Pred. Bagus	7	2	77.78%
Pred. Buruk	8	33	80.49%
Class recall	46.67%	94.29%	

Berdasarkan tabel diatas, dapat dilihat hasil akurasi yang diperoleh sebesar 80%, lebih detailnya algoritma memprediksi 33 item memiliki kualitas buruk dan memang kualitas sarang walet tersebut buruk (TP). Selanjutnya algoritma ini memprediksi 7 item memiliki kualitas yang bagus dan sesuai dengan kualitas sarang walet tersebut bagus (TN), Diikuti dengan item bagus berjumlah 8 akan tetapi algoritma memprediksi item tersebut buruk (FP) yang merupakan Type I Error. Terakhir 2 item sarang walet buruk tetapi di prediksi bagus (FN) yang merupakan jenis Type II Error.

Pada kasus kali ini, Type I error (FP) cukup berbahaya karna dapat merugikan penjual disaat item bagus di prediksi oleh algoritma sebagai item buruk. Lebih lanjut untuk mengukur kinerja model yang dibuat dapat dilihat pada tabel diatas untuk performance metrics yang digunakan yaitu: accuracy, precission dan recall. Precission sendiri digunakan untuk menggambarkan tingkat akurasi antara data dengan hasil prediksi oleh model dan



algoritma (TP). Sedangkan recall digunakan untuk menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi.

Tabel 12. Hasil Precision Dari Evaluasi Dengan Confusion Matrix

precision: 80.49% (positive class: Buruk)			
	True Bagus	True Buruk	Class Precision
Pred. Bagus	7	2	77.78%
Pred. Buruk	8	33	80.49%
Class recall	46.67%	94.29%	

Tabel 13. Hasil Recall Dari Evaluasi Dengan Confusion Matrix

recall: 94.29% (positive class: Buruk)			
	True Bagus	True Buruk	Class Precision
Pred. Bagus	7	2	77.78%
Pred. Buruk	8	33	80.49%
Class recall	46.67%	94.29%	

Sesuai dengan nilai precision maka dapat disimpulkan bahwa dengan 33 item diprediksi benar kualitas buruk dan 8 ternyata berkualitas baik diperoleh 80.49%. Serta nilai recall dengan 33 item nya diprediksi benar kualitas buruk dan 2 ternyata bernilai bagus diperoleh 94.29%. Kita dapat menggunakan accuracy sebagai acuan performa algoritma jika dataset memiliki jumlah FN dan FP yang mendekati simetris. Jika tidak, dapat melanjutkan menggunakan Perhitungan F1 Score sebagai acuan performa algoritma. Hal ini dapat dilihat pada tabel dibawah:

Tabel 14. Hasil F1 Scire Dari Evaluasi Dengan Confusion Matrix

F_measure: 86.84% (positive class: Buruk)			
	True Bagus	True Buruk	Class Precision
Pred. Bagus	7	2	77.78%
Pred. Buruk	8	33	80.49%
Class recall	46.67%	94.29%	

Berdasarkan tabel diatas dapat dilihat bahwa semua hasil evaluasi yang diperoleh memiliki nilai tinggi baik dari hasil akurasi, nilai precision maupun nilai recall. Bahkan dalam penelitian ini juga dilakukan perhitungan F1 score yang juga memperoleh persentase yang tinggi. Dengan melihat hasil tersebut dapat disimpulkan bahwa klasifikasi berhasil dilakukan.

4. KESIMPULAN

Penelitian ini dilakukan untuk mengklasifikasikan hasil panen sarang walet dengan menggunakan teknik data mining dengan metode atau algoritma K-Nearest Neighbor. Klasifikasi sarang walet dibuat dengan atribut kebersihan, warna, ukuran, bentuk dan masa panen yang masuk kedalam kategori "Bagus" atau "Buruk" dan menggunakan 49 data sampel dan 1 data sampel baru atau pembandingan. Hasil klasifikasi K-neraest neighbor yang diperoleh dengan perhitungan WEKA yaitu 45 data mendapat point akurasi sebesar 90% dengan nilai $k = 7$ yang artinya sampel data yang digunakan terklasifikasi dengan benar. Hal tersebut didukung dengan hasil evaluasi yang diperoleh dari confusion matrix menggunakan rapidminer dengan tingkat akurasi 80%. Diperkuat dengan performance metrics lain nya yaitu Precision dengan 80.49%, Nilai Recall dengan 94.29 dan karna FP dan FN tidak mendekati simetris maka dibuktikan juga dengan nilai F yaitu 86.84%. ini menggambarkan algoritma ini memiliki tingkat akurasi yang cukup tinggi dan klasifikasi kualitas sarang burung walet yang dilakukan juga memiliki tingkat keabsahan yang tinggi.

REFERENCES

- [1] A. P. P. Wicaksono, N. K. Kusmayati, dan T. Kurniawan, "Pengaruh Usaha Sarang Burung Walet Terhadap Pendapatan Masyarakat Di Kota Surabaya," *Jurnal Riset Bisnis dan Ekonomi*, vol. 4, no. 1, 2023.
- [2] F. E. Y. Kha, T. Uda, S. Rohaetin, R. Alexandro, dan D. Erang, "Manfaat Sosial Ekonomi Budidaya Sarang Burung Walet Bagi Masyarakat," *Jurnal Ilmu Ekonomi & Sosial*, vol. 12, no. 2, 2021.
- [3] Husdi dan M. G. Sayoga, "PROTOTYPE ALAT PEMANTAUAN DAN PENGENDALIAN SUHU RUANGAN SARANG WALET BERBASIS NODEMCU," *Jurnal Sistem Informasi Dan Teknik Komputer*, vol. 8, no. 1, 2023.
- [4] R. Rakhmadi, A. Hadiawan, D. Muhammad, dan S. Zahratun, "Potensi Ekspor Sarang Burung Walet Provinsi Lampung," *Jurnal Hubungan Internasional Indonesia*, vol. 4, no. 1, 2022, [Daring]. Tersedia pada: <http://jhii.fisip.unila.ac.id/ojs/index.php/jhii>
- [5] D. Damayanti, "Implementasi Algoritma C4.5 Prediksi Produksi Komoditas Tanaman Perkebunan Berdasarkan Luas Lahan," *Jurnal Terapan Informatika Nusantara*, vol. 2, no. 10, hlm. 571–579, Mar 2022, doi: 10.47065/tin.v2i10.1026.



- [6] A. Lili, Suhada, dan S. Widodo, "Pengelompokan Hasil Panen Kelapa Sawit Dalam Produksi Per Blok Menggunakan Algoritma K-Means," *Journal of Machine Learning and Data Analytics (MALDA)*, vol. 01, no. 01, 2022.
- [7] E. P. W. Mandala dan D. E. Putri, *Data Mining Asosiasi dan Klasterisasi Produk pada Toko Retail*. Solok: Penerbit Insan Cendekia Mandiri, 2022.
- [8] F. Maulana, M. Orisa, dan H. Z. Zahro, "KLASIFIKASI DATA PRODUK MEBEL ANEKA JAYA MENGGUNAKAN METODE K-NEAREST NEIGHBOR BERBASIS WEB," *Jurnal Mahasiswa Teknik Informatika*, vol. 5, no. 2, 2021.
- [9] M. Wibowo dan R. Ramadhani, "Perbandingan Metode Klasifikasi Data Mining Untuk Rekomendasi Tanaman Pangan," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 3, hlm. 913, Jul 2021, doi: 10.30865/mib.v5i3.3086.
- [10] B. Basuki, A. Nazir, S. K. Gusti, L. Handayani, dan I. Iskandar, "Klasifikasi Tingkat Keberhasilan Produksi Ayam Broiler di Riau Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 3, hlm. 493, Mar 2023, doi: 10.30865/json.v4i3.5665.
- [11] A. D. W. Sumari, P. I. Mawarni, dan A. R. Syulistyo, "Klasifikasi Mutu Telur Burung Puyuh Berdasarkan Warna dan Tekstur Menggunakan Metode K-Nearest Neighbor (KNN) dan Fusi Informasi," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 5, hlm. 1019, Okt 2021, doi: 10.25126/jtiik.2021854393.
- [12] E. P. W. Mandala, D. E. Putri, dan R. Permana, "Penerapan Data Mining untuk Klasifikasi Hasil Panen Jamur Tiram Menggunakan Algoritma K-Nearest Neighbors," *Jurnal Media Informatika Budidarma*, vol. 7, no. 1, 2023, doi: 10.30865/mib.v7i1.5252.
- [13] M. Jumarlis, Mirfan, dan A. R. Manga, "Classification of Coffee Bean Defects Using Gray-Level Co-Occurrence Matrix and K-Nearest Neighbor," *ILKOM Jurnal Ilmiah*, vol. 14, no. 1, hlm. 1–9, Apr 2022, doi: 10.33096/ilkom.v14i1.910.1-9.
- [14] D. N. Aini, B. Oktavianti, M. J. Husain, D. A. Sabillah, S. T. Rizaldi, dan Mustakim, "Seleksi Fitur untuk Prediksi Hasil Produksi Agrikultur pada Algoritma K-Nearest Neighbor (KNN)," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 1, hlm. 140, Sep 2022, doi: 10.30865/json.v4i1.4813.
- [15] S. P. Nabila, N. Ulinuha, dan A. Yusuf, "MODEL PREDIKSI KELULUSAN TEPAT WAKTU DENGAN METODE FUZZY C-MEANS DAN K-NEAREST NEIGHBORS MENGGUNAKAN DATA REGISTRASI MAHASISWA," *Jurnal Ilmiah NERO*, vol. 6, no. 1, 2021.
- [16] A. Bode, Z. Y. Lamasigi, dan I. C. R. Drajana, "The K-Nearest Neighbor algorithm using Forward Selection and Backward Elimination in predicting the Student's Satisfaction Level of University Ichsan Gorontalo toward Online Lectures during the COVID-19 Pandemic," *ILKOM Jurnal Ilmiah*, vol. 15, no. 1, 2023.
- [17] U. Erdiansyah, A. I. Lubis, dan K. Erwansyah, "Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kulit," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, hlm. 208, Jan 2022, doi: 10.30865/mib.v6i1.3373.
- [18] P. R. Prayogo dan P. H. Susilo, "Sistem Pendukung Keputusan dalam Menentukan Kualitas Sarang Burung Walet Terbaik Menggunakan Metode Simple Additive Weighting (SAW)," *Insearch (Information System Research) Journal*, vol. 2, no. 2, 2022.
- [19] R. N. Sucihati, Usman, dan R. D. Kantari, "ANALISIS PENDAPATAN DAN KELAYAKAN BUDIDAYA SARANG BURUNG WALET DI KECAMATAN LUNYUK," *Jurnal Ekonomi dan Bisnis*, vol. 8, no. 2, 2020.
- [20] I. A. Fausy, E. Meiyani, dan S. Amin, "Tingkat Kesejahteraan Masyarakat Dan Usaha Sarang Burung Walet Di Desa Belopa Kabupaten Luwu," *Journal Socius Education (JSE)*, vol. 1, no. 1, 2023, doi: 10.0505/jse.v%vi%i.416.
- [21] T. Arifin dan D. Ariesta, "PREDIKSI PENYAKIT GINJAL KRONIS MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER BERBASIS PARTICLE SWARM OPTIMIZATION," *Jurnal Tekno Insentif*, vol. 13, no. 1, hlm. 26–30, Apr 2019, doi: 10.36787/jti.v13i1.97.
- [22] D. Normawati dan S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, hlm. 697–711, 2021.
- [23] D. Putra dan A. Wibowo, "Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes," dalam *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, 2020, hlm. 84–92.