

## ANALISIS PENGOLAHAN EKSTRAKSI FITUR CITRA UNTUK KLASIFIKASI JENIS APEL MENGGUNAKAN SCIKIT-LEARN DENGAN ALGORITMA K-NEAREST NEIGHBOR

Robi Wariyanto Abdullah<sup>\*1</sup>, Rajnaparamitha Kusumastuti<sup>2</sup>, Handoko<sup>3</sup>

<sup>1,2,3</sup>Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Amikom Surakarta, Kabupaten Sukoharjo

Email: <sup>1</sup>robiwa@dosen.amikomsolo.ac.id, <sup>2</sup>rajna@dosen.amikomsolo.ac.id,

<sup>3</sup>Handoko@dosen.amikomsolo.ac.id

\*Penulis Korespondensi

(Naskah masuk: 31 Agustus 2024, diterima untuk diterbitkan: 11 Februari 2025)

### Abstrak

Jenis Apel di Indonesia makin beragam seiring dengan perkembangan teknologi dibidang perkebunan.apel yang berkembang pesat. Penggunaan teknologi pengolahan citra dan pembelajaran mesin telah membuka peluang baru dalam mengatasi tantangan klasifikasi objek berbasis citra, termasuk dalam mengidentifikasi jenis buah apel.Penggunaan metode K-Nearest Neighbor sudah banyak terbukti dalam klasifikasi berbagai jenis data termasuk citra. Penelitian yang akan dilakukan akan mencoba melakukan analisa dan mengklasifikasikan jenis apel dengan membaca extrasi fitur citra menggunakan library scikit-learn dalam bahasa pemrograman Python dengan pendekatan algoritma K-Nearest Neighbor. Dataset yang akan diteli dalam klasifikasi digunakan 7 jenis apel yaitu Apple Braeburn, Apple Crimson Snow, Apple Golden, Apple Granny Smith, Apple Red , Apple Red Delicious, Apple Red Yellow. Dataset training yang digunakan dalam penelitian sebanyak 16404 citra, sedangkan data testing sebanyak 2134 citra apel. Proses klasifikasi dilakukan dengan membandingkan extrasi fitur dari apel yang belum diketahui varietasnya dengan fitur-fitur dari apel yang telah diklasifikasikan sebelumnya.extrasi fitur yang akan dilakukan yaitu akan membandingkan dari hasil extrasi fitur HVS, histogram dan RGB yang dipilih dengan nilai k genap. Hasil penelitian yang dilakukan menghasilkan akurasi tertinggi sebesar 96,9 %, dengan nilai k=2 dan kriteria perhitungan jarak menggunakan ekstraksi fitur menggunakan RGB,HSV dan histogram. Penelitian ini diharapkan mampu memberikan potensi penggunaan teknik pengolahan citra dalam mendukung identifikasi jenis apel secara otomatis, yang relevan dalam industri pertanian dan pengolahan makanan. Penelitian selanjutnya diharapkan dapat dilakukan dengan membandingkan metode algoritma untuk klasifikasi yang lain serta memberikan dataset image dengan pencahayaan yang berbeda dengan menambah beberapa jenis apel dan kombinasi parameter yang lebih banyak lagi agar dapat meningkatkan output hasil penelitian yang sudah dilakukan.

**Kata kunci:** *Klasifikasi, K-Nearest Neighbor, Pembelajaran Mesin, Python, scikit-learn*

### **IMAGE FEATURE EXTRACTION PROCESSING ANALYSIS FOR APPLE TYPE CLASSIFICATION USING SCIKIT-LEARN WITH THE K-NEAREST NEIGHBOR ALGORITHM**

#### **Abstract**

*Apple types in Indonesia are increasingly diverse along with the rapid development of technology in the field of apple plantations. The use of image processing and machine learning technology has opened up new opportunities in overcoming the challenges of image-based object classification, including in identifying apple types. The use of the K-Nearest Neighbor method has been widely proven in the classification of various types of data including images. The research that will be conducted will try to analyze and classify apple types by reading image feature extraction using the scikit-learn library in the Python programming language with the K-Nearest Neighbor algorithm approach. The dataset that will be studied in the classification uses 7 types of apples, namely Apple Braeburn, Apple Crimson Snow, Apple Golden, Apple Granny Smith, Apple Red, Apple Red Delicious, Apple Red Yellow. The training dataset used in the study was 16404 images, while the testing data was 2134 apple images. The classification process is carried out by comparing feature extraction from apples whose varieties are unknown with features from apples that have been previously classified. The feature extraction that will be carried out is to compare and from the results of the HVS, histogram and RGB feature extraction selected with an even k value. The results of the research conducted produced the highest accuracy of 96.9%, with a value of k = 2 and the*

*distance calculation criteria using feature extraction using RGB, HSV and histograms. This research is expected to provide the potential for using image processing techniques to support automatic identification of apple types, which are relevant in the agricultural and food processing industries. Further research is expected to be carried out by comparing algorithm methods for other classifications and providing image datasets with different lighting by adding several types of apples and more parameter combinations in order to increase the output of the research results that have been carried out.*

**Keywords:** *Classification, K-Nearest Neighbor, Machine Learning, Python, scikit-learn*

## 1. PENDAHULUAN

Di Indonesia, kebun apel biasanya terletak di dataran tinggi yang beriklim sejuk, seperti daerah pegunungan di beberapa propinsi seperti Jawa Barat, Jawa Timur, Jawa Tengah, Sumatera Barat, dan Sumatera Utara. Daerah yang terkenal dengan kebun apel di Indonesia diantaranya kabupaten batu, Jawa timur, kabupaten malang jawa timur, kabupaten wonosobo, jawa tengah, kabupaten solok Sumatra barat.

Meskipun Indonesia tidak memiliki iklim yang ideal untuk pertumbuhan apel, beberapa daerah tertentu dengan kondisi topografi yang tepat mampu menghasilkan apel dengan kualitas yang baik dengan menghasilkan beberapa jenis apel dalam kebun mereka. Apel banyak digemari masyarakat selain banyak nutrisi dan vitamin, apel memiliki berbagai jenis yang dapat dilihat dari bentuk, warna, ukuran, Jenis-jenis apel yang umumnya dikenal masyarakat dan mudah ditemui dipasaran yaitu Red Delicious, Apple Fuji Jingle, Apple Golden Delicious, Gala, Grannysmith, Manalagi, dan Malang dan mungkin masih banyak lagi jenis apel yang lain.

Dengan banyaknya jenis buah apel dalam perkebunan apel, bagi masyarakat umum masih belum begitu memahami ciri khas masing-masing jenis apel termasuk dalam jenis apel yang mana, sehingga dalam penelitian ini mencoba untuk menganalisis dari jenis apel melalui pengolahan ekstrasi fitur citra dari apel untuk mengklasifikasikan jenis apel tersebut.

Salah satu tahap dalam penelitian untuk mengklasifikasikan jenis apel yaitu dengan menerapkan teknik supervised learning yang merupakan suatu metode pendekatan pada machine learning yang dapat memperoleh informasi data yang membutuhkan data training pada setiap jenis dan label yang bersesuaian untuk mempelajari dari data yang baru / data testing (Aherwadi and Mittal, 2022). Dalam penelitian yang akan dilakukan untuk mengklasifikasikan jenis apel akan digunakan pemrograman python sehingga diperlukan library scikit-learn dengan memanfaatkan google colab sebagai tool untuk menjalankan kode program python. Library scikit-learn merupakan library Python yang dikembangkan untuk memberikan kemudahan melakukan analisis kode program machine learning dalam Python yang dikembangkan oleh beberapa kontributor yang sudah banyak digunakan di dunia industri maupun akademisi (Fahmi, 2023), salah satu

implementasi yaitu untuk mengklasifikasikan jenis apel dengan menggunakan algoritma k-nearest neighbor.

Penelitian ini diharapkan menghasilkan informasi dan kontribusi yang bermanfaat bagi perkebunan apel dalam mengklasifikasikan jenis apel. sehingga dengan menerapkan klasifikasi dengan memanfaatkan machine learning dapat membantu masyarakat sekitar dapat mudah memahami, mengenal jenis apel yang ada dipasaran dan diharapkan mampu memberikan potensi penggunaan teknik pengolahan citra dalam mendukung identifikasi jenis apel secara otomatis, yang relevan dalam industri pertanian dan pengolahan makanan.

Penelitian sebelumnya telah mengungkapkan berbagai aspek yang relevan dengan topik klasifikasi dengan K-Nearest Neighbors. salah satu penelitian yang dilakukan oleh sugiyono tentang pengolahan citra untuk klasifikasi jenis manga dengan K-Nearest Neighbors menggunakan aplikasi matlab dengan data training sebanyak 390 dan data uji sebanyak 240 object menghasilkan akurasi pengujian sebesar 61%. (Sugiyono and Ruswandi, 2022). Penelitian yang serupa yang dilakukan oleh sugiyono (Sugiyono and Ruswandi, 2022) tentang klasifikasi jenis pisang menggunakan K-Nearest Neighbors dengan data training sebanyak 50 dari 3 jenis pisang sehingga data training yang digunakan yaitu 150 data dan dengan data uji 120 data gambar menghasilkan akurasi sebesar 99,16%

Penelitian yang mengklasifikasikan kualitas apel dari warna dan bentuk dengan K-Nearest Neighbors yang dilakukan oleh Cindy Suryanti (Suryanti *et al.*, 2024) memiliki 2 class yaitu baik dan buruk dengan data training 74 object gambar dan data testing sebesar 43 object gambar menghasilkan akurasi sebesar 86,04 % dengan menggunakan k=3, dan menghasilkan akurasi sebesar 88,37 % dengan k=7 dan menghasilkan akurasi sebesar 83,37% pada k=9, sedangkan pada penelitian (Hassan *et al.*, 2023) melakukan penelitian untuk 13 jenis apel dengan menghasilkan akurasi sebesar 92%.

Penelitian yang dilakukan oleh Wahyu Saputro (Saputro and Sumantri, 2022) dalam menggunakan metode KNN dan implementasi citra digital untuk mengklasifikasikan jenis buah anggur dengan 6 class buah anggur menghasilkan akurasi sebesar 80,98% dengan data training sebanyak 3419 dan data testing sebanyak 984 dengan software pengolahan menggunakan matlab.

Penelitian yang dilakukan oleh Moh. Arie Hasan (Hasan, 2023) dalam mengklasifikasikan jenis apel Braeburn dan Crimson Snow. dengan KNN digunakan Principal Component Analysis untuk mengolah segmentasi citra dan menghasilkan akurasi sebesar 91,67% dengan k yang digunakan yaitu k=3 dengan jumlah data training yang digunakan sebanyak 30 object gambar untuk jenis apel Braeburn dan 30 object gambar jenis apel Crimson Snow dan dengan menggunakan data testing sebanyak 12 object

Penelitian yang dilakukan oleh (Sarimole and Fadillah, 2022) tentang klasifikasi buah jambu dengan KNN menggunakan fitur HVS, pada penelitian tersebut menggunakan tool matlab, sedangkan pada penelitian yang dilakukan yaitu akan mengkombinasikan fitur HVS, RGB dan histogram dengan menggunakan tool google colab dengan memanfaatkan scikit-learn pada python.

Penelitian yang dilakukan oleh (Hasan, 2023) untuk mengklasifikasikan jenis apel menggunakan fitur PCA menghasilkan akurasi sebesar 91,16%. Pengolahan analisis data yang digunakan menggunakan matlab. Sedangkan pada penelitian yang akan dilakukan yaitu akan menggunakan Scikit-learn untuk ekstraksi fitur dengan menggunakan HSV, RGB dan histogram. Beberapa penelitian yang serupa untuk klasifikasi suatu object menggunakan tool rata-rata yaitu menggunakan matlab seperti pada penelitian (Nababan, Khairi and Harahap, 2022), (Adenugraha, Arinal and Mulyana, 2022)

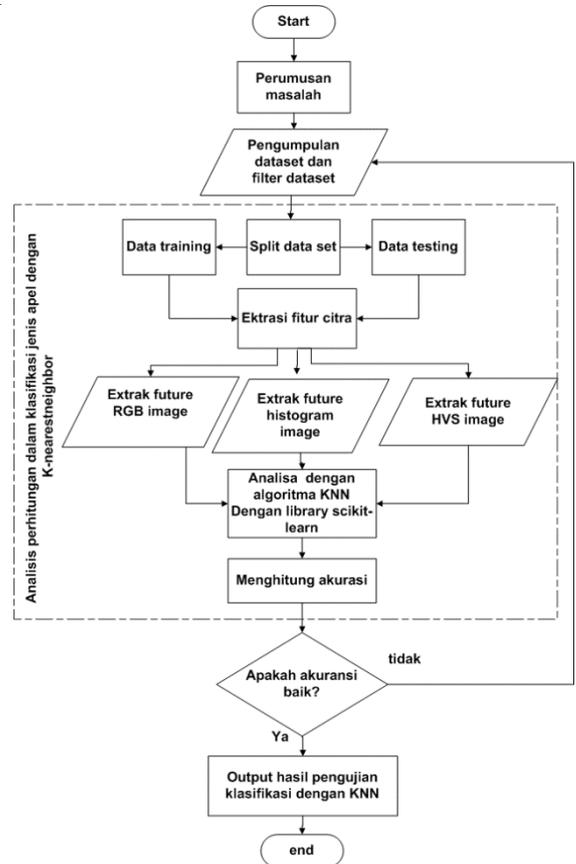
Penelitian yang menggunakan scikit-learn dilakukan oleh peneliti arpit patidar (Chakravorty, 2024) untuk mengidentifikasi penyakit dan menyortir buah apel . dalam penggunaan scikit learn peneliti menggunakan algoritma CNN untuk mengidentifikasi penyakit buah apel dengan menghasilkan akurasi sebesar 95,37%.

Berdasarkan ulasan berbagai studi terdahulu, penelitian ini bertujuan untuk melaksanakan eksperimen dengan judul "Analisis Pengolahan Ekstraksi Fitur Citra untuk klasifikasi Jenis Apel Menggunakan scikit-learn dengan Algoritma K-nearest neighbor". Pemilihan library scikit-learn dengan algoritma K-nearest Neighbor dilakukan untuk melanjutkan dan memperluas penelitian yang telah ada, dengan menguji penerapannya pada dataset yang berbeda dengan tool dan bahasa pemrograman python dengan klasifikasi pembelajaran 7 jenis apel untuk mengevaluasi apakah dapat dicapai tingkat akurasi yang signifikan lebih tinggi dibandingkan dengan penggunaan dataset yang memiliki class dibawah 7 dengan library pengolahan image dan tool analisis yang berbeda pada penelitian sebelumnya..

**2. METODE PENELITIAN**

Perangkat keras yang digunakan adalah laptop dengan spesifikasi prosesor AMD A9 3,1 GH dan perangkat lunak yang digunakan untuk menganalisa dan klasifikasi data yaitu dengan menggunakan google colab yang nantinya akan dihubungkan ke

google drive untuk menjalankan program python untuk membaca dan mengambil data training dan data testing yang sudah diupload di google drive untuk dilakukan analisis dan klasifikasi data. Alur penelitian



Gambar 2. Alur Proses Penelitian

Tahap awal dari penelitian yaitu melakukan pengamatan masalah tentang klasifikasi apel dengan melihat literature penelitian sebelumnya , kemudian dilakukan perumusan masalah tentang klasifikasi apel . Tahap yang dilakukan selanjutnya yaitu mengumpulkan dataset training dan data testing serta menfilter data yang akan dilakukan analisa dan klasifikasi.

Perkembangan teknologi memudahkan kita dalam pengambilan data secara digital, sumber data telah menjadi bagian integral dalam berbagai bidang. Salah satu tempat untuk mendapatkan data bagi para peneliti, ilmuwan data, dan praktisi machine learning untuk mendapatkan dataset adalah Kaggle.

Proses pengambilan data dari kaggle yaitu dengan mencari dataset buah apel yang akan digunakan dalam analisis pada platform kaggle kemudian data yang telah ditemukan sesuai dengan dataset yang akan diteliti maka dataset tersebut di download dari kaggle. hasil pencarian yang didapat penulis menemukan dataset yaitu fruits-360 dengan versi: 2020.05.18.0 yang berisi 90483 gambar buah dan sayuran dimana data tersebut terdiri dari 67692 gambar dan data testing 22688 gambar dengan 131

class yang terbentuk dari buah dan sayuran . Dataset yang telah diperoleh dari kaggle maka data tersebut hanya diambil pada buah apple yang nantinya akan digunakan untuk klasifikasi jenis apple. Data training dan data testing yang sudah difilter maka data tersebut diupload di google drive yang nantinya akan digunakan untuk analisis klasifikasi jenis apel.

Pada bagian ekstraksi fitur citra, gambar 2 menunjukkan tiga jenis ekstraksi yang dilakukan, yaitu: Ekstrak fitur RGB image, Ekstrak fitur histogram image, Ekstrak fitur HVS image. Ekstraksi fitur RGB bertujuan untuk mengambil nilai rata-rata dari setiap komponen warna (Red, Green, Blue) dalam gambar. Hasil rata-rata ini akan digunakan sebagai vektor fitur untuk klasifikasi. Histogram warna digunakan untuk menangkap distribusi intensitas dari setiap warna dalam citra. Ini memberikan lebih banyak informasi mengenai variasi warna yang ada. Ekstraksi fitur HVS mengonversi citra dari RGB ke ruang warna HVS, yang lebih stabil terhadap perubahan pencahayaan. Nilai rata-rata dari masing-masing komponen (Hue, Saturation, Value) akan digunakan sebagai fitur. Setelah fitur RGB

Histogram, dan HVS diekstraksi peneliti akan melakukan 3 kali pengujian dengan menghitung nilai akurasi dari kombinasi setiap hasil ekstraksi tersebut dengan memanfaatkan library scikit-learn. pengujian pertama akan dilakukan dengan menghitung akurasi dengan menggunakan hasil fitur ekstraksi RGB dan histogram, pengujian yang kedua yaitu dengan melakukan perhitungan akurasi dari hasil ekstraksi fitur HVS dan histogram, sementara pengujian yang ketiga akan melakukan perhitungan akurasi dengan menggunakan hasil kombinasi ketiga ekstraksi fitur yaitu RGB, HVS dan histogram. Ekstraksi fitur dilakukan menggunakan library OpenCV untuk manipulasi warna dan histogram, sementara scikit-learn mengelola pembagian dataset, pelatihan model, dan evaluasi akurasi.

Pada tahap analisa dengan algoritma KNN dengan library scikit-learn, scikit-learn memainkan peran utama dalam mengimplementasikan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan gambar berdasarkan fitur-fitur yang sudah diekstraksi (fitur RGB, histogram, dan HVS). K-Nearest Neighbor (KNN) adalah algoritma pembelajaran mesin berbasis instance-based learning, di mana keputusan klasifikasi untuk data baru dilakukan dengan cara mencari sejumlah tetangga terdekat yang sudah diketahui kelasnya dalam ruang fitur. KNN merupakan method yang simple dalam mengklasifikasikan suatu object (Nababan, Khairi and Harahap, 2022) , (Profile, 2024). Algoritma ini termasuk dalam kategori supervised learning, yang berarti algoritma ini membutuhkan data yang telah terklasifikasi dalam suatu class (Suryanti *et al.*, 2024). Pembentukan Dataset untuk KNN Setelah fitur RGB, histogram, dan HVS diekstraksi dari setiap gambar, fitur-fitur ini digabungkan untuk dilakukan pengujian menjadi satu vektor fitur gabungan. Setiap

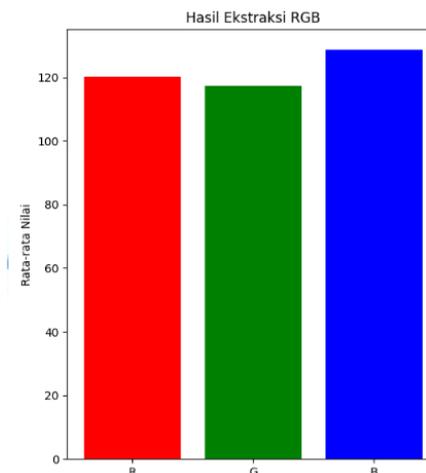
vektor fitur ini memiliki label yang menunjukkan kelas gambar / jenis apel. Library Scikit-learn membutuhkan format data berupa array untuk pembentukan class jenis apel tersebut. Peneliti membuat class diambil dari nama folder untuk tiap jenis apel yang sudah terklasifikasi sebagai dataset training yang nantinya akan diolah oleh library scikit-learn.

Setelah melakukan pelatihan model K-nearest neighbor menggunakan data training, peneliti perlu mengevaluasi kinerja model tersebut menggunakan data testing. Scikit-learn menyediakan berbagai fungsi dan metrik yang dapat digunakan untuk mengevaluasi kinerja model klasifikasi. , peneliti dapat menggunakan fungsi `accuracy_score` dari scikit-learn untuk menghitung akurasi. Fungsi `accuracy_score` dari scikit-learn menerima dua argumen: `y_test` (label asli) dan `y_pred` (label hasil prediksi model). Fungsi ini akan mengembalikan nilai akurasi sebagai sebuah nilai desimal. Confusion matrix adalah salah satu metode evaluasi yang lebih rinci dibandingkan akurasi keseluruhan karena memberikan gambaran tentang bagaimana model klasifikasi berperilaku pada tiap kelas dalam data (Hamzah, Susanti and Lestari, 2024). Confusion matrix menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas, memungkinkan kita melihat apakah model cenderung salah klasifikasi pada kelas tertentu. Scikit-learn menyediakan metode `confusion_matrix` untuk menghitung dan menampilkan hasil akurasi yang lebih detail

### 3. HASIL DAN PEMBAHASAN

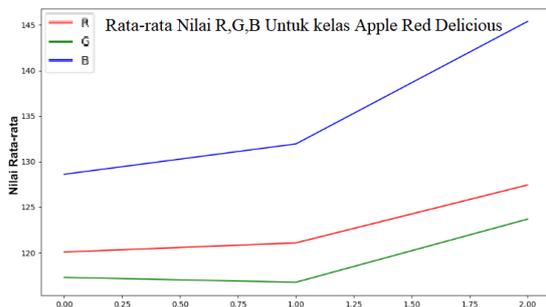
#### 3.1 Ekstraksi Fitur RGB

Dataset yang sudah didapat dari kaggle, baik data training maupun data testing maka hal yang pertama dilakukan peneliti yaitu mengextract fitur RGB, sebagai contoh dari hasil output yang didapat dari ekstraksi fitur image beberapa class yang akan diuji maka peneliti mengambil contoh pada jenis apel red delicious, Hasil ekstraksi dari apel delicious dapat dilihat pada gambar 3.



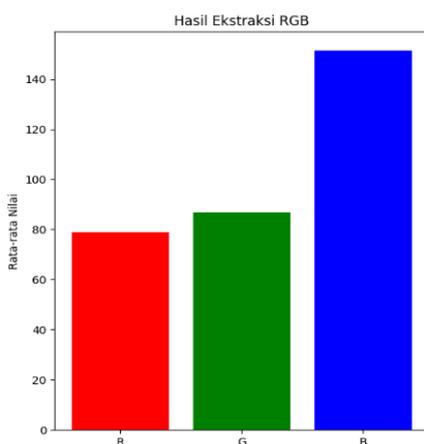
Gambar 3. Ekstraksi RGB jenis buah apel Red Delicious

Diagram yang terlihat pada gambar 3 merupakan representasi hasil ekstraksi dari salah satu gambar yang peneliti ambil sebagai dari hasil ekstraksi pada apel red delicious. Hasil ekstraksi RGB sendiri merupakan warna yang dihasilkan oleh cahaya dengan panjang gelombang tertentu dalam spektrum yang terlihat. Pada diagram di atas terlihat bahwa pada apel red delicious menghasilkan ekstraksi warna R sebesar 120, G menghasilkan 110 dan B menghasilkan 130. Sedangkan untuk image lain yang memiliki jenis yang sama menghasilkan tidak jauh berbeda dalam satu class yang sama yaitu pada gambar apel red delicious.



Gambar 4. Rata-rata nilai RGB untuk jenis apel red delicious

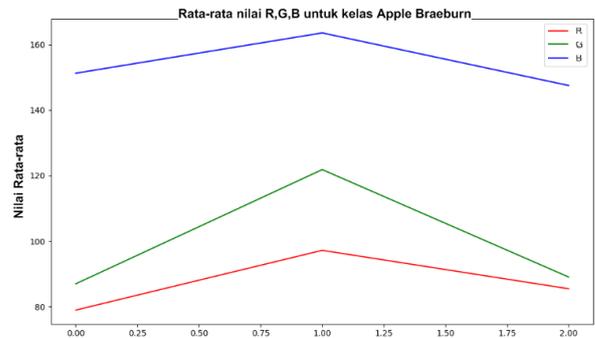
Diagram yang terlihat pada gambar 4 menunjukkan rata-rata hasil ekstraksi pada jenis apel red delicious. Jenis apel red delicious memiliki nilai RGB yang sama-sama tinggi baik nilai R, G maupun B. Hal tersebut dikarenakan pada apel red delicious memiliki merah gelap sehingga dihasilkan warna pixel merah yang cukup padat. Nilai dari B / blue memiliki nilai yang paling tinggi yaitu rata-rata warna biru dalam apel terdeteksi 128-145.



Gambar 5. Ekstraksi RGB jenis buah apel braeburn

Diagram yang terlihat pada gambar 5 merepresentasikan dari hasil ekstraksi salah satu gambar yang peneliti ambil dari jenis apel braeburn. Hasil menunjukkan bahwa tingkat warna yang dihasilkan pada B/ warna biru juga termasuk tinggi seperti pada apel red delicious akan tetapi tingkat

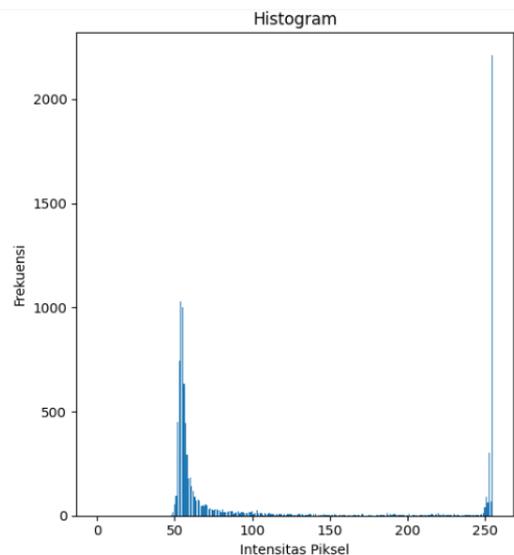
warna pixel merah dan pixel hijau pada jenis apel braeburn cenderung sedikit dimana tingkat warna R memiliki nilai 70 sedangkan pada G memiliki nilai 90 dan B 145. Pada image yang memiliki jenis yang sama yaitu apel braeburn setelah dilakukan ekstraksi image memiliki hasil rgb yang tidak jauh beda pada image dari class yang sama yaitu pada apel braeburn.



Gambar 6. Rata-rata nilai RGB apel Braeburn

Diagram yang terlihat pada gambar 6 merepresentasikan bahwa gambar dataset pada class / jenis apel braeburn memiliki rata-rata nilai RGB yang mewakili disetiap gambar menghasilkan nilai B cukup tinggi akan tetapi nilai R dan G tidak terlalu banyak memberikan kontribusi nilai pixel warna tersebut.

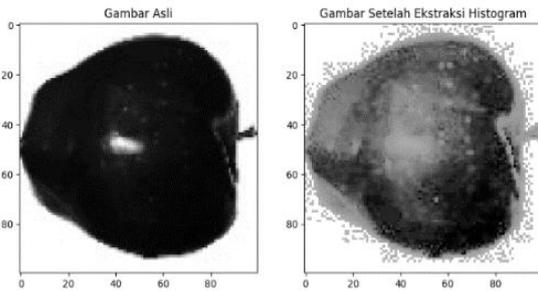
### 3.2 Ekstrasi Fitur Histogram



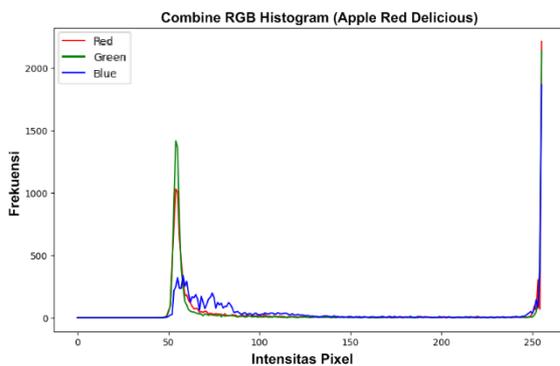
Gambar 7. Ekstrasi Histogram apel red delicious

Hasil Ekstrasi fitur gambar histogram pada salah satu gambar jenis apel red delicious yaitu terlihat pada gambar 7, diagram tersebut menunjukkan bahwa pada sumbu horizontal menunjukkan rentang nilai piksel pada gambar, sedangkan pada sumbu vertical menggambarkan jumlah / frekuensi kemunculan piksel pada nilai tertentu. Semakin tinggi puncak pada diagram maka semakin banyak piksel dalam

gambar yang memiliki nilai kecerahan atau kegelapan yang sesuai. Hasil gambar setelah dilakukan ekstrasi histogram dapat dilihat pada gambar 8

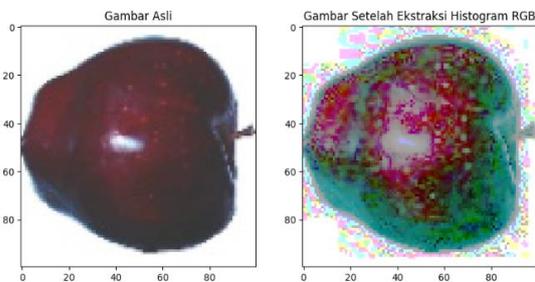


Gambar 8. Hasil Gambar dengan histogram apel red delicious



Gambar 9. Ekstrasi Histogram RGB apel red delicious

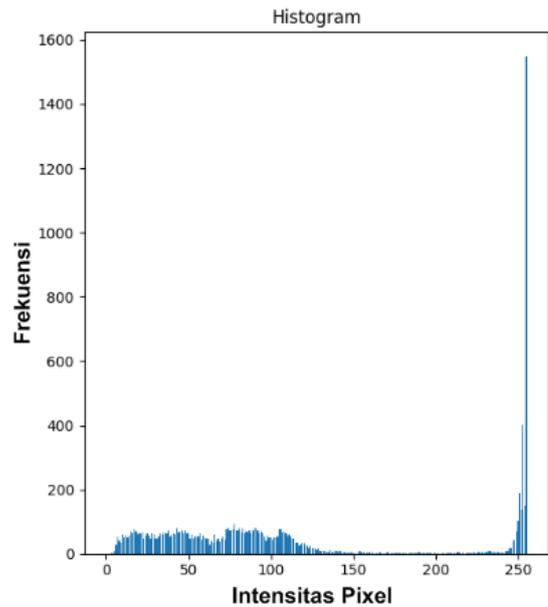
Setelah melakukan ekstrasi dengan histogram pada salah satu gambar jenis apel red delicious maka peneliti melakukan percobaan untuk mengekstrasi histogram dengan kombinasi ekstrasi RGB, hasil ekstrasi dari histogram RGB dapat dilihat pada gambar 9. Pada diagram tersebut menunjukkan bahwa puncak pada sumbu biru menunjukkan seberapa banyak piksel dalam gambar memiliki nilai intensitas biru tertentu. Semakin tinggi puncaknya, semakin banyak piksel yang memiliki nilai intensitas biru tersebut, begitu juga untuk nilai pada puncak sumbu hijau dan merah. Hasil gambar setelah dilakukan ekstrasi histogram RGB dapat dilihat seperti pada gambar 10.



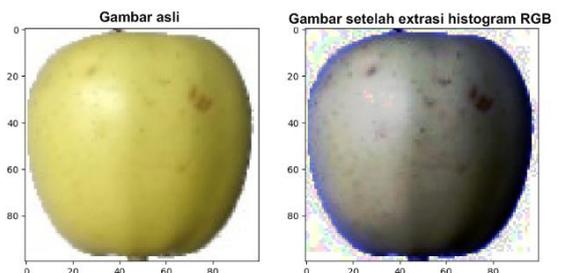
Gambar 10. Ekstrasi Histogram RGB apel red delicious

Peneliti melakukan percobaan untuk menunjukkan hasil ekstrasi gambar pada salah satu gambar pada jenis apel yang berbeda yaitu pada jenis apel golden. Hasil ekstrasi fitur golden dapat dilihat pada gambar 11. Diagram histogram pada apel golden

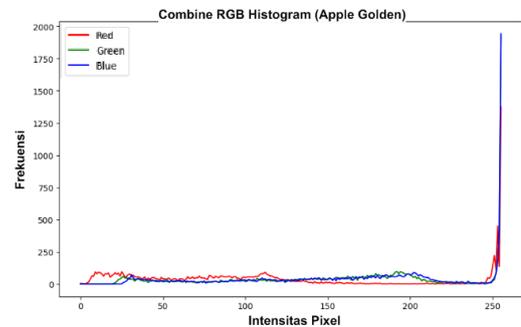
sangat berbeda dengan jenis apel pada red delicious, pada apel red delicious intensitas piksel yang padat hanya pada bagian tertentu sedangkan pada apel golden intensitas piksel yang diperoleh tidak begitu padat pikselnya akan tetapi disetiap bagian memiliki jumlah piksel yang merata. Hasil gambar pada jenis apel yang sama yaitu pada apel golden menghasilkan diagram grafik yang mendekati/serupa dari percobaan hasil ekstrasi pada gambar jenis apel yang sama.



Gambar 11. Ekstrasi Histogram apel golden



Gambar 12. Hasil Gambar Ekstrasi Histogram apel golden



Gambar 13. Ekstrasi histogram RGB apel golden

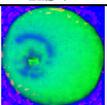
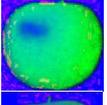
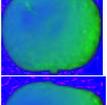
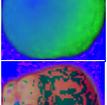
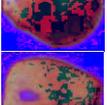
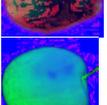
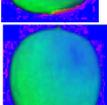
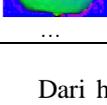
Setelah melakukan ekstrasi dengan histogram pada salah satu gambar jenis apel red delicious dan golden maka peneliti akan melakukan percobaan juga pada salah satu gambar jenis apel golden untuk mengekstrasi histogram dengan kombinasi ekstrasi RGB, hasil diagram dari ekstrasi histogram RGB dapat

dilihat pada gambar 13. Pada diagram tersebut menunjukkan bahwa gambar memiliki intensitas yang berbeda dengan apel red delicious setelah dilakukan ekstrasi fitur gambar histogram yang dikombinasikan dengan RGB, dari gambar 13 merepresentasikan jumlah pixel tidak begitu padat baik untuk nilai R, G maupun B, peneliti juga mencoba mengextrasi gambar lain pada jenis apel yang sama dan hasil yang didapatkan memiliki kedekatan tidak begitu jauh berbeda pada gambar hasil pengujian yang gambar pertama pada jenis apel golden. Gambar dari salah satu gambar percobaan untuk mengetahui hasil ekstrasi histogram RGB dari jenis apel golden dapat dilihat pada gambar 12

### 3.3 Ekstrasi Fitur HVS

Setelah melakukan ekstrasi RGB dan histogram peneliti mencoba juga dalam mengextrasi gambar dengan HVS. Tabel 1 menunjukkan hasil dari ekstrasi gambar yang peneliti ambil yaitu hanya mengambil 4 contoh jenis apel dengan mengambil 2 contoh gambar disetiap class/ jenis apel untuk dapat diketahui nilai HVS dari gambar tersebut.

Tabel 1. Ekstrasi Citra HVS

Image HSV	Class	Hue	Saturation	value
	Apple Braeburn	20.1519	148.4037	147.7053
	Apple Braeburn	45.9084	138.7154	161.9241
	Apple Golden	28.4164	102.0584	174.7466
	Apple Golden	28.0569	101.4149	173.2489
	Apple Red Delicious	99.6294	50.0119	134.6507
	Apple Red Delicious	98.2912	56.9749	147.0876
	Apple red yellow	28.7829	94.4806	165.5695
	Apple red yellow	21.5729	118.6919	185.0267
...	...	...	...	...

Dari hasil yang terdapat pada Tabel 1 terlihat bahwa nilai H, V, S disetiap gambar pada class yang sama memiliki nilai kedekatan nilai HVS tersebut. Sebagai contoh pada apel goldel pada image 1 menghasilkan nilai Hue=28, Value=102, Saturation=174 dan pada image 2 pada jenis apel

golden tidak begitu berbeda nilai yang dihasilkan dikarenakan dari segi warna, bentuk memiliki kemiripan yang hampir sama dibanding dengan jenis apel yang lain. Setelah mendapatkan nilai HVS peneliti juga menampilkan hasil image HVS tersebut.

### 3.4 Perhitungan Algoritma KNN

Penggunaan library scikit-learn merupakan tool utama yang digunakan dalam perhitungan algoritma dengan K-nearest neighbor. Untuk dapat menggunakan library tersebut maka penting untuk mengimport terlebih dahulu pada model yang dibuat dengan google colab seperti pada gambar 14.

```
import os
import cv2
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report, accuracy_score
```

Gambar 14. Import library scikit-learn python

Import os digunakan untuk berinteraksi dengan sistem operasi salah satu penggunaannya yaitu untuk membaca atau menulis file pada direktori tertentu. import cv2 OpenCV digunakan untuk pemrosesan gambar, dalam penelitian ini opencv digunakan untuk membantu mengextrasi fitur image yang digunakan yaitu dengan ekstrasi RGB, HVS dan Histogram. import numpy as np: library yang digunakan untuk komputasi numerik yang berguna dalam mengelola array dan operasi matriks pada algoritma KNN.

Fungsi train\_test\_split digunakan untuk membagi data menjadi set pelatihan dan set pengujian, yang memungkinkan evaluasi kinerja model. fungsi KNeighborsClassifier dari library scikit-learn merupakan model K-Nearest Neighbors (KNN) yang digunakan dalam klasifikasi jenis apel. Algoritma KNN bekerja dengan mencari tetangga terdekat dari data baru dan memprediksi kelas berdasarkan mayoritas kelas dari tetangga terdekat tersebut. classification\_report digunakan untuk menghasilkan laporan kinerja model, yang mencakup metrik seperti precision, recall, dan F1-score untuk setiap kelas. accuracy\_score digunakan untuk menghitung akurasi model, yaitu proporsi prediksi yang benar dibandingkan dengan jumlah total prediksi.

```
k = 5
model = KNeighborsClassifier(n_neighbors=k)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
```

Gambar 15. Penggunaan skikit learn dengan metode KNN

Parameter k menentukan jumlah tetangga terdekat yang akan digunakan dalam klasifikasi oleh algoritma K-Nearest Neighbors (KNN). Dalam hal ini, setiap data baru akan diklasifikasikan berdasarkan mayoritas kelas dari 5 tetangga terdekat. pada model = KNeighborsClassifier(n\_neighbors=k) kode ini

membuat model KNN dengan menggunakan kelas KNeighborsClassifier dari scikit-learn, dengan jumlah tetangga (n\_neighbors) diset ke nilai k, yaitu 5. Model KNN bekerja dengan mencari k tetangga terdekat dari titik data yang akan diklasifikasikan. model KNN dilatih menggunakan data pelatihan X\_train dan labelnya y\_train. Metode .fit() akan membuat model berdasarkan data pelatihan sehingga dapat memprediksi label untuk data baru. Setelah model dilatih, fungsi predict digunakan pada data pengujian X\_test untuk menghasilkan prediksi label, yang disimpan dalam variabel y\_pred.accuracy\_score menghitung akurasi model berdasarkan perbandingan antara label yang sebenarnya y\_test dengan prediksi model y\_pred. Akurasi menunjukkan proporsi prediksi yang benar dibandingkan dengan total prediksi. Fungsi classification\_report digunakan untuk menghasilkan laporan detail tentang kinerja model. Laporan ini mencakup metrik seperti precision, recall, F1-score, dan support untuk setiap kelas yang ada.

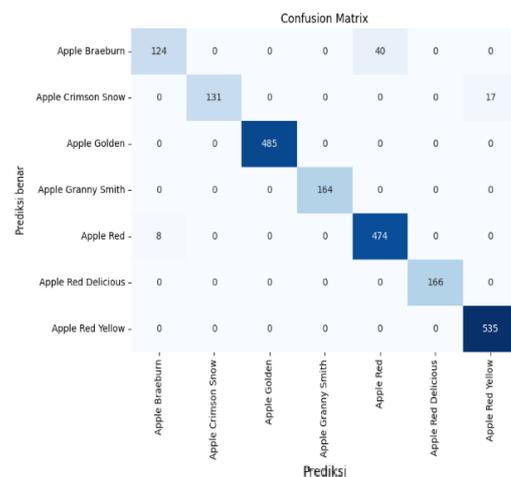
Setelah melakukan ekstraksi citra yang akan digunakan yaitu dengan ekstraksi RGB,HVS dan Histogram maka untuk pengujian yang pertama peneliti akan melakukan pengujian terlebih dahulu untuk menggunakan 2 fitur ekstraksi yaitu RGB dan histogram untuk mengetahui tingkat akurasi dan waktu yang dibutuhkan dalam klasifikasi jenis apel dari data testing terhadap data training dengan perhitungan algoritma KNN. Hasil pengujian ketika menggunakan 2 fitur tersebut menghasilkan tingkat akurasi tertinggi yaitu pada nilai k=2 dengan tingkat akurasi sebesar 83,8% dengan waktu yang dibutuhkan rata-rata 2 menit . Hasil akurasi dengan nilai k yang lain dapat dilihat pada table 2.

Setelah melakukan pengujian yang pertama maka peneliti mencoba untuk mengkombinasikan jenis ekstraksi yang berbeda yaitu dengan ekstraksi HVS dengan ekstraksi Histogram. Hasil pengujian dengan menggunakan fitur ekstraksi tersebut menghasilkan tingkat akurasi sebesar 95,1% dengan nilai k=2 dan waktu analisis rata-rata 2 menit 21 detik.. Dari hasil pengujian yang kedua terlihat bahwa tingkat akurasi mengalami kenaikan sebesar 13,49%. Hasil akurasi pada nilai k lainnya dapat dilihat pada tabel 2.

Pengujian berikutnya yaitu peneliti akan menggunakan 3 ekstraksi sekaligus yaitu RGB,HVS dan histogram. Hasil penelitian dengan menggunakan ketiga ekstraksi fitur citra tersebut menghasilkan akurasi sebesar 96,9% rata-rata waktu yang digunakan untuk menganalisis sebesar 2 menit 35 detik. Pada table 4 terlihat bahwa hasil tingkat akurasi dengan k yang lebih tinggi menunjukkan akurasi semakin rendah. Dan dari beberapa pengujian yang telah dilakukan dengan 2 fitur ekstraksi dengan 3 fitur ekstraksi tingkat akurasi yang tertinggi yaitu dengan menggunakan 3 ekstraksi fitur citra yaitu dengan ekstraksi fitur citra RGB,HVS dan histogram.

Tabel 2. Perhitungan Akurasi dengan Pengujian kombinasi ekstraksi fitur

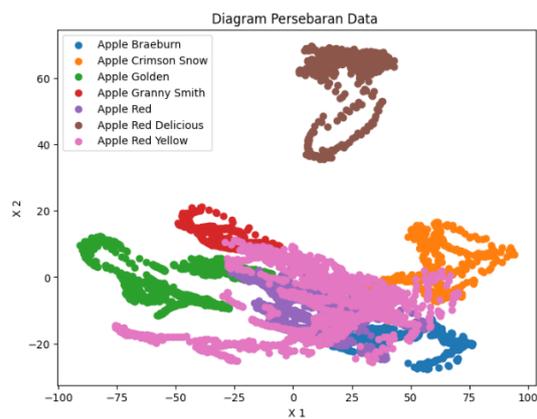
Nilai k	Nilai akurasi		
	RGB dan Histogram	HVS dan histogram	RGB,HVS dan histogram
2	0,838	0,951	0,969
4	0,831	0,936	0,952
6	0,839	0,934	0,951
8	0,836	0,925	0,950
10	0,834	0,921	0,950
12	0,833	0,914	0,945
14	0,829	0,911	0,944
16	0,824	0,909	0,942
18	0,817	0,905	0,940



Gambar 16. Confusion Matrix

Diagram confusion matrik yang terlihat pada gambar 16 merepresentasikan bahwa dalam memprediksi klasifikasi apel braburn sebanyak 124 gambar apel braeburn diprediksi benar masuk dalam class apel braeburn sedangkan sebanyak 40 gambar apel Braeburn diprediksi kurang tepat yang seharusnya masuk dalam class braeburn akan tetapi diprediksi sebagai apel red , sedangkan sebanyak 166 gambar apel red diprediksi benar , dan 8 gambar yang seharusnya termasuk dalam class apel red diprediksi masuk sebagai class apel braeburn, sedangkan untuk class apel yang lain sudah diprediksi tepat sesuai dengan class atau jenis apel yang sebenarnya. Dari kesalahan tersebut disebabkan karena kurangnya varian gambar pada apel braeburn dan apel red yang menyebabkan tingkat akurasi pengujian yang tertinggi mencapai 96,9%.

Diagram yang ditunjukkan pada gambar 15 merepresentasikan kedekatan hasil perhitungan tiap jenis apel setelah dilakukan perhitungan dengan algoritma K-nearest neighbor dengan menghitung jarak Euclidean distance data testing terhadap data training. Hasil dari ke tujuh jenis apel sebagian besar memiliki titik kedekatan sesuai dengan masing-masing jenis apel.



Gambar 15. Diagram Persebaran Jenis Apel

#### 4. KESIMPULAN

Hasil pengujian penelitian yang dilakukan dalam mengklasifikasikan jenis buah apel dengan 7 class/ jenis apel diantaranya Apple Braeburn, Apple Crimson Snow, Apple Golden, Apple Granny Smith, Apple Red, Apple Red Delicious, Apple Red Yellow menghasilkan nilai akurasi tertinggi sebesar 96,9% dengan nilai  $k=2$  dan kriteria ekstrasi yang digunakan yaitu dengan memberikan pengujian terhadap kombinasi jenis ekstrasi fitur citra menggunakan ekstrasi RGB, HVS dan histogram dengan rata-rata membutuhkan waktu 2 menit untuk analisa dataset data training sebesar 6404 citra apel dan data testing sebesar 2134 citra apel.

Untuk memperkuat hasil yang didapat maka diharapkan penelitian berikutnya diharapkan dapat membandingkan dengan algoritma klasifikasi lain seperti Support Vector Machine (SVM), Random Forest, atau Convolutional Neural Network (CNN) untuk membandingkan tingkat akurasi dengan K-Nearest Neighbor (KNN) selain itu dapat ditambahkan fitur citra selain RGB, HVS, dan histogram, seperti tekstur (contohnya Gray Level Co-occurrence Matrix atau GLCM) dan bentuk, untuk memperbaiki kemampuan klasifikasi dalam membedakan jenis apel yang memiliki warna serupa. selain menambahkan variasi jenis apel atau bahkan menambah jumlah gambar pada setiap jenis apel dalam dataset. diharapkan juga dalam penelitian berikutnya diuji coba dengan kondisi cahaya yang berbeda pada pengambilan gambar apel dapat dilakukan untuk melihat apakah metode ekstraksi fitur yang digunakan tahan terhadap perubahan pencahayaan.

#### DAFTAR PUSTAKA

- ADENUGRAHA, S.P., ARINAL, V. AND MULYANA, D.I. 2022. Klasifikasi Kematangan Buah Pisang Ambon Menggunakan Metode KNN dan PCA Berdasarkan Citra RGB dan HSV. *Jurnal Media Informatika Budidarma*, 6(1), p. 9.
- AHERWADI, N. AND MITTAL, U. 2022. Fruit

Quality Identification Using Image Processing, Machine Learning, and Deep Learning: a Review. *Advances and Applications in Mathematical Sciences*, 21(5), pp. 2645–2660.

- CHAKRAVORTY, A. 2024. Using Machine Learning to Identify Diseases and Perform Sorting in Apple Fruit. (July).
- FAHMI, M.N. 2023. Implementasi Mechine Learning menggunakan Python Library: Scikit-Learn (Supervised dan Unsupervised Learning). *Sains Data Jurnal Studi Matematika dan Teknologi*, 1(2), pp. 87–96.
- HAMZAH, A., SUSANTI, E. AND LESTARI, R.M. 2024. Klasifikasi Kematangan Buah Alpukat Mentega Berdasarkan Fitur Warna Menggunakan Support Vektor Machine (Svm), 6(1), pp. 108–120.
- HASAN, M.A. 2023. Classification of Apple Types Using Principal Component Analysis and K-Nearest Neighbor. *International Journal of Information System Technology and Data Science (IJIST-DAS)*, 1(1), pp. 15–22.
- HASSAN, N.M.H. ET AL. 2023. Quality of performance evaluation of ten machine learning algorithms in classifying thirteen types of apple fruits, (April), pp. 102–109.
- NABABAN, A.A., KHAIRI, M. AND HARAHAP, B.S. 2022. Implementation of K-Nearest Neighbors (KNN) algorithm in classification of data water quality. *Jurnal Mantik*, 6(1), pp. 30–35.
- PROFILE, S.E.E. 2024. Classification of Apple Quality Using XGBoost Machine Learning Model XGBoost Makine Öğrenimi Modelini Kullanarak Apple Kalitesinin Sınıflandırılması. *4th International Conference on Innovative Academic Studies Classification of Apple Quality Using XGBoost Machine Learning Model XGBoost Makine Öğrenimi Modelini Kullanarak Apple Kalitesinin Sınıflandırılması*, (March).
- SAPUTRO, W. & SUMANTRI, D.B. 2022. Implementasi Citra Digital Dalam Klasifikasi Jenis Buah Anggur Dengan Algoritma K-Nearest Neighbors (KNN) Dan Data Augmentasi. *INTECOMS: Journal of Information Technology and Computer Science*, 5(2), pp. 248–253.
- SARIMOLE, F.M. AND FADILLAH, M.I. 2022. Classification of Guarantee Fruit Murability Based on Hsv Image With K-Nearest Neighbor. *Journal of Applied Engineering and Technological Science*, 4(1), pp. 48–57.
- SUGIYONO, S. AND RUSWANDI, M. 2022. Pemodelan Pengolahan Citra untuk Klasifikasi Jenis Buah Pisang Menggunakan Metode KNN. *Jurnal ...*, 4(5), pp. 823–833.
- SURYANTI, C. ET AL. (2024) ‘Klasifikasi Kualitas

Buah Apel Berdasarkan Warna dan Bentuk  
Menggunakan Metode KNN, *Generation*

*Journal*, 8(1), pp. 2580–4952.