



## PERANCANGAN APLIKASI PENENTUAN KUALITAS SAYURAN BERDASARKAN WARNA MENGGUNAKAN DATA MINING

Purwa Hasan Putra<sup>1</sup>, Muhammad Syahputra Novelan<sup>2</sup>,  
Muhammad Rizki<sup>3</sup>

Fakultas Sains dan Teknologi, Universitas Pembangunan Panca Budi  
[purwahasanputra@dosen.pancabudi.ac.id](mailto:purwahasanputra@dosen.pancabudi.ac.id)

### ABSTRACT

Dalam penelitian ini akan dilakukan proses menerapkan metod *K-Nearest Neighbor* (KNN) merupakan salah satu metode klasifikasi terhadap sekumpulan data yang berdasarkan mayoritas dari kategori dan tujuannya untuk mengklasifikasikan obyek baru berdasarkan atribut dan *sample* dari *training data*. Sehingga target *output* yang diinginkan mendekati ketepatan dalam melakukan pengujian pembelajaran. Hasil pengujian metode *K-Nearest Neighbor* Adapun terlihat bahwa dari nilai *K* 1 hingga 10 yang di uji persentase dari hasil analisis metode *K-NN* lebih tinggi dari hasil analisis metode *K-NN*. Dan dari nilai *K* yang telah di ujikan nilai *K* 2 dan nilai *K* 9 memiliki persentasi terbesar sehingga keakuratannya juga semakin tepat. Adapun dari hasil pengujian metode *K-Nearest Neighbor* dalam klasifikasi data. Adapun pengujian penulis menggunakan variasi nilai *K* *K-Nearest Neighbor* 3,4,5,6,7,8,9. Memiliki persentase akurasi yang sangat baik di dibandingkan hanya *K-NN*. Hasil pengujian menunjukan metode *K-Nearest Neighbor* dalam klasifikasi data memiliki akurasi persentase yang baik ketika menggunakan data *random*. Persentase variasi nilai *K* *K-Nearest Neighbor* 3,4,5,6,7,8,9 memiliki persentase 100 % .

### PENDAHULUAN

Proses klasifikasi citra ini mengacu pada metode kecerdasan buatan yang memfokuskan pada pembelajaran mesin (*machine learning*). Banyak metode lain dalam pembelajaran mesin (*machine learning*) yang digunakan untuk proses klasifikasi diantaranya *K-Nearest Neighbor* dan *Naïve Bayes Classifier*. Klasifikasi merupakan pengelompokan suatu objek ke dalam kelas-kelas berdasarkan ciri-ciri persamaan maupun perbedaan.

Adege *et al.* (2018) dalam penelitiannya tentang *Indoor Localization* menggunakan metode *K-Nearest Neighbor* (k-NN) dan *Backpropagation* mendapatkan hasil bahwa metode k-NN menghasilkan akurasi yang lebih baik dibandingkan metode *Backpropagation*. Sedangkan Jaafar *et al.* (2016) dalam penelitiannya menggunakan metode k-NN untuk mengklasifikasi basis data gambar biometrik berbasis tangan (*hand-based biometric*) yang merupakan basis data sidik jari (*fingerprint*) dan vena jari (*finger vein*), serta melakukan optimasi pada metode kNN untuk mendapatkan persentase yang lebih baik.

Dengan menggunakan metode *K-Nearest Neighbor* (KNN) merupakan salah satu metode klasifikasi terhadap sekumpulan data yang berdasarkan mayoritas dari kategori dan tujuannya untuk mengklasifikasikan obyek baru berdasarkan atribut dan *sample* dari *training data*. Sehingga target *output* yang diinginkan mendekati ketepatan dalam melakukan pengujian pembelajaran.

### TINJAUAN PUSTAKA

Machine learning memungkinkan dalam klasifikasi data, aplikasi ini mengenali pola dalam data baik dengan pelatihan atau tanpa pelatihan. Di klasifikasi data disebut clustering dalam machine learning. Beberapa contoh algoritma pengelompokan termasuk *K-Means*, *Farthest-First* Maksimalisasi-Ekspektasi (EM), dan lainnya (Ion C. Freeman, Ashley J. et al, 2018).



Machine Learning merupakan cabang ilmu artificial intelligence yang memiliki konsep bahwa komputer sebagai mesin memiliki kemampuan untuk melakukan adaptasi terhadap lingkungan yang baru dan mampu mendeteksi pola dari fakta yang ada. Definisi pembelajaran yang dilakukan mesin adalah ketika mesin dari pengalaman E terhadap tugas T dan mengukur peningkatan kinerja P, jika kinerja tugas T diukur oleh kinerja P, meningkatkan pengalaman E.

### **Klasifikasi Data**

Klasifikasi data adalah proses pengurutan dan pengelompokan data ke dalam berbagai jenis, bentuk atau kelas berbeda lainnya. Klasifikasi data memungkinkan pemisahan dan klasifikasi data sesuai dengan persyaratan kumpulan data untuk berbagai tujuan bisnis atau pribadi. Ini terutama merupakan proses manajemen data.

Analisis kelompok sebagai suatu metode untuk melakukan klasifikasi data menjadi beberapa kelompok dengan menggunakan metode pengukuran ukuran asosiasi, sehingga data yang sama berada dalam satu kelompok dan data yang memiliki perbedaan yang besar diletakkan dalam kelompok data lainnya. Masukan untuk sistem analisis kelompok adalah sebuah data set dan kesamaan ukuran antara kedua data tersebut. Sedangkan hasil dari analisis kelompok adalah sejumlah kelompok yang membentuk sebuah partisi atau struktur partisi dari kumpulan data dan deskripsi umum dari setiap kelompok, dimana hal ini sangat penting untuk analisis yang lebih dalam pada karakteristik yang terdapat pada data tersebut.

Pengelompokan data harus menggunakan sebuah pendekatan untuk mencari kesamaan dalam data sehingga mampu menempatkan data ke dalam kelompok-kelompok yang tepat. Pengelompokan data akan membagi kumpulan data ke dalam beberapa kelompok dimana kesamaan dalam sebuah kelompok lebih besar jika dibandingkan dengan kelompok-kelompok lainnya (Gorunescu, 2011).

Terdapat dua metode pembelajaran yang tersedia pada model klasifikasi yaitu :

- a. Eager learning adalah proses pembelajaran pada data latih secara intensif sehingga model dapat melakukan prediksi ke label kelas output dengan benar. Beberapa metode bersifat eager learning antara lain, Neural Network, Bayesian, decision tree, Support Vector Machine.
- b. Lazy learning adalah proses pembelajaran tanpa melakukan pelatihan dan hanya menyimpan nilai data latih untuk digunakan pada proses prediksi. Beberapa metode bersifat lazy learning antara lain: K-Nearest Neighbor, Regresi Linear, Fuzzy K-Nearest Neighbor.

Proses klasifikasi pada machine learning memiliki empat komponen (Saputra, 2018) yaitu :

1. Class

Variabel dependen tetamu harus dalam bentuk yang mewakili label yang dimiliki oleh objek.

2. Energi

Variabel independen yang diwakili oleh atribut karakteristik data. Misalnya, gaji, kehadiran, merokok, tekanan darah.

3. Dataset Pelatihan

Satu set data yang memiliki kedua nilai komponen diatas yang digunakan untuk menentukan kelas yang sesuai berdasarkan energi.

4. Menguji dataset

Satu set data baru yang akan diklasifikasi dengan model yang telah dibuat dan akan dievaluasi pada proses akurasi klasifikasi.

Pada proses klasifikasi, sebelum melakukan prediksi, perlu dilakukan proses pembelajaran terlebih dahulu. Proses pembelajaran tersebut memerlukan data. Data yang

diperlukan pada saat proses klasifikasi terdiri atas dua jenis, yaitu :

- Data latih atau data training adalah data yang digunakan pada proses pembelajaran dalam proses klasifikasi.
- Data uji atau data testing adalah data yang digunakan pada proses prediksi dalam proses klasifikasi.

### Euclidean Distance

Jarak Euclidean hanya merupakan penjumlahan dari perbedaan intensitas piksel-bijaksana dan, akibatnya, deformasi kecil dapat menghasilkan jarak Euclidean yang besar. Makalah ini mengusulkan jarak Euclidean baru, yang kami sebut Gambar Jarak Euclidean (IMED). Tidak seperti yang tradisional, IMED mempertimbangkan hubungan spasial piksel. Berdasarkan tiga properti yang (boleh dibilang) metrik gambar yang dapat secara intuitif memenuhi, kami menunjukkan bahwa IMED adalah satu-satunya jarak Euclidean yang memiliki properti ini. Jarak Euclidean Distance adalah jarak antar titik dalam garis lurus. Metode jarak ini menggunakan teorema Pythagoras. Dan merupakan perhitungan jarak yang paling sering digunakan dalam proses machine learning (Viriyavisuthisakul et al, 2015). Rumus Euclidean Distance adalah hasil dari akar kuadrat perbedaan dua buah vektor.

## METODOLOGI PENELITIAN

### Pengumpulan Data

*Data yang dikumpulkan untuk penelitian ini yaitu file image (citra) dengan format Portable Network Graphics (PNG) yang diperoleh dengan menggunakan kamera digital. Citra yang menjadi inputan adalah citra buah pepaya. Data sampel yang digunakan adalah sebanyak 3 data pada masing-masing sampel citra, dengan masing-masing memiliki 3 atribut yaitu red, green, blue. Dataset tersebut adalah hasil ekstraksi citra yang akan menjadi sumber data untuk klasifikasi citra buah pepaya dengan metode K-Nearest Neighbor.*

### Pemrosesan Citra Buah

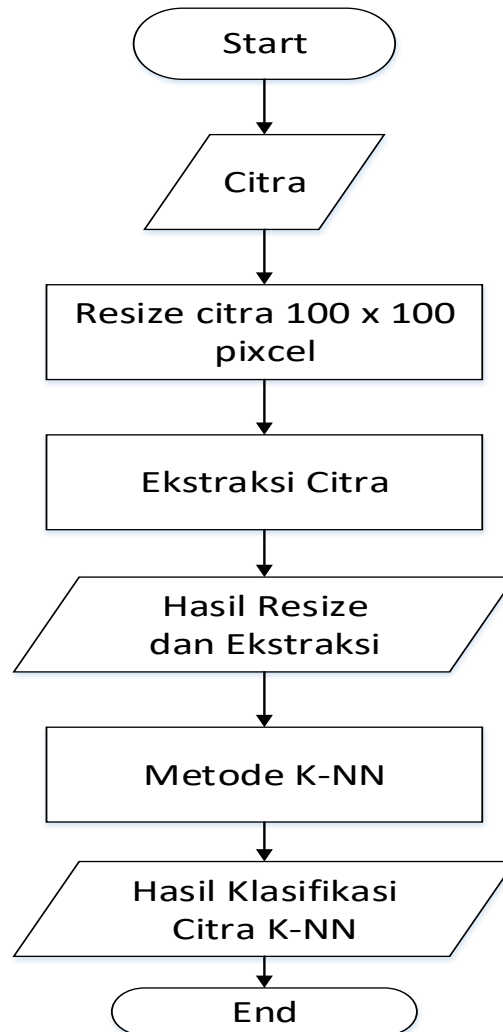
Pada pemrosesan citra (image processing) akan dilakukan menjadi 2 tahap yaitu resize citra dan proses ekstraksi yang juga diuraikan dalam bentuk pseudo-code. Berikut sebuah sampel citra pepaya yang akan diidentifikasi tingkat kematangannya. Citra yang digunakan dalam sampel ini adalah citra pepaya asli (original) yang akan di-resize menjadi ukuran 100 x 100 piksel, tujuannya agar proses klasifikasi lebih mudah dilakukan dikarenakan keseragaman dalam ukuran citra. Citra original dan citra yang telah di-resize ditunjukkan pada Gambar 1.



Gambar 1. Citra Pepaya



## Flowchart



Gambar 2. Flowchart Penelitian

## HASIL DAN PEMBAHASAN

### Pembahasan

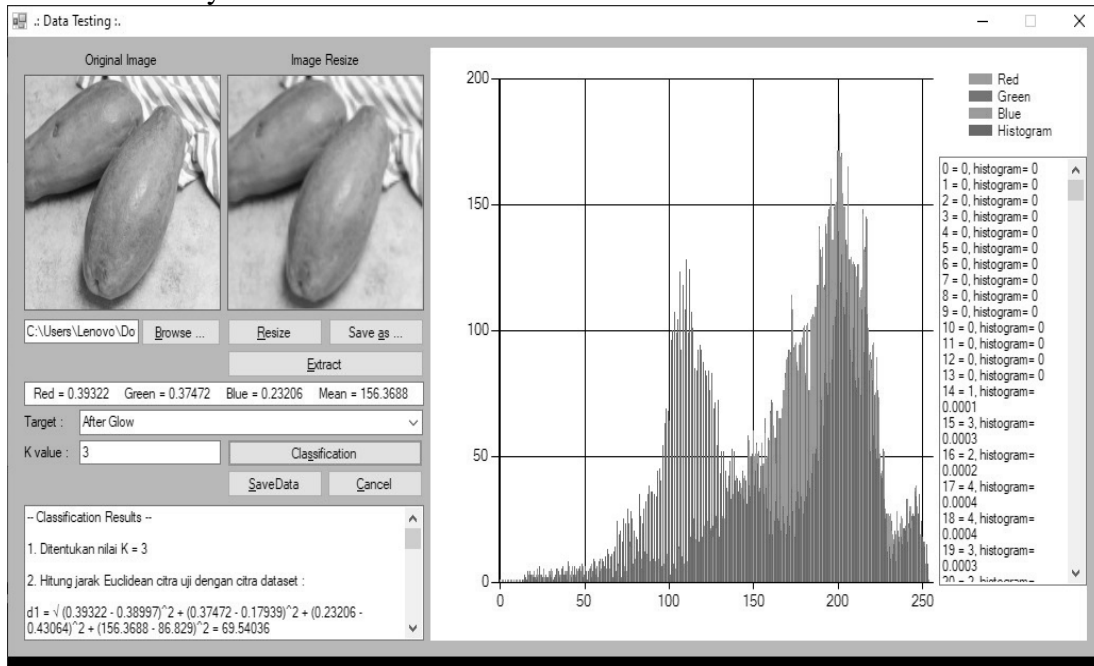
Dalam penelitian ini akan dilakukan proses penentuan jenis data buah pepaya Bougenville atau yang biasa disebut dengan buah pepaya kertas ini merupakan tanaman hias yang eksistensinya sudah cukup populer di kalangan masyarakat dan tersebar luas di berbagai wilayah di Indonesia. Data yang dikumpulkan untuk penelitian ini yaitu file image (citra) dengan format Portable Network Graphics (PNG) yang diperoleh dengan menggunakan kamera digital. Citra yang menjadi inputan adalah citra **Kualitas Buah Pepaya**. Data sampel yang digunakan adalah sebanyak 3 data pada masing-masing sampel citra, dengan masing-masing memiliki 3 atribut yaitu red, green, blue. Dataset tersebut adalah hasil ekstraksi citra yang akan menjadi sumber data untuk klasifikasi citra buah dengan metode K-Nearest Neighbor.

### Proses Klasifikasi Citra

Aplikasi yang dirancang dapat diimplementasikan dalam klasifikasi citra buah pepaya berdasarkan tingkat kematangannya dengan metode yang telah dimasukkan ke dalam

algoritmanya. Adapun analisis metode K-Nearest Neighbor dalam klasifikasi citra sayuran yang telah diimplementasikan ke dalam aplikasi dapat dijabarkan sebagai berikut

Sebelum melakukan pengujian, terlebih dahulu harus disiapkan data latih yang berupadataset hasil ekstraksi citra yang disimpan ke dalam database. Gambar berikut menunjukkan aplikasi yang dikembangkan untuk mengekstraksi citra latih dan menyimpan hasil ekstraksinya ke dalam database.



**Gambar 3. Ekstraksi Citra Latih**

Pada Gambar diatas dapat dilihat proses pengolahan citra untuk mendapatkan dataset. Terdapat proses resize, extract, dan save data. Proses resize digunakan untuk mengubah ukuran atau resolusi citra menjadi seragam, yaitu dalam kasus ini 100 x 100 piksel. Setelah citra di-resize, selanjutnya diekstrak untuk mendapatkan nilai atribut red, green, blue, hue, saturation, value, mean, variance, skewness, kurtosis, dan energy.

Pada aplikasi juga ditampilkan grafik nilai red, green, blue, dan histogram pada citra yang nantinya nilai-nilai tersebut diproses untuk mendapatkan nilai atribut yang telah ditentukan. Hasil ekstraksi tersebut selanjutnya akan disimpan ke dalam database dengan memberikan label target yang sesuai dengan tingkat warna buah pepaya yang nantinya akan digunakan sebagai data uji.

### Hasil Pembahasan

Pada penelitian ini penulis menganalisis pengujian dengan variasi nilai (K) K-NN. Dari hasil analisis menunjukkan Pada pengujian ini penulis juga menganalisis variasi nilai K dari buah pepaya. Adapun ditunjukkan sebagai berikut.

Dalam pengujian ini menggunakan 30 data uji dengan 4 atribut dan 3 species dalam klasifikasi data.

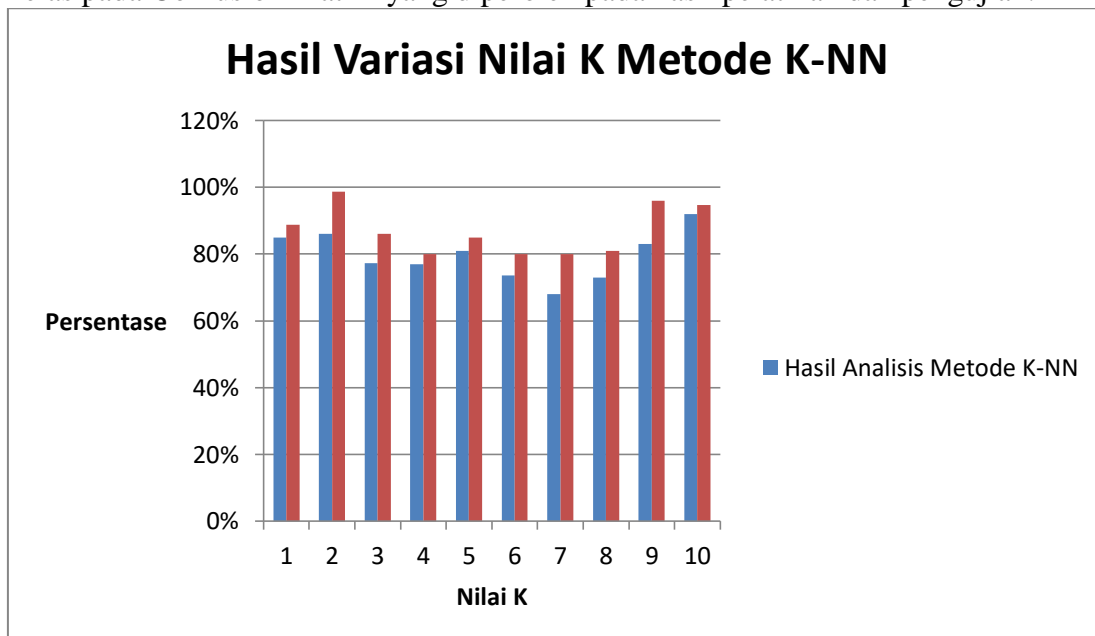
**Tabel 1. Hasil Variasi Nilai K Metode K-NN**

Data Buah Pepaya	Jumlah Nilai (K) K-NN	Hasil Analisis Metode K-NN
	3, 5, 7, 8, 9	85%
	3, 5, 7, 8, 9	86%

	3, 5, 7, 8, 9	77.3%
	3, 5, 7, 8, 9	77%
	3, 5, 7, 8, 9	81%
	3, 5, 7, 8, 9	73.6%
	3, 5, 7, 8, 9	68%
	3, 5, 7, 8, 9	73%
	3, 5, 7, 8, 9	83%
	3, 5, 7, 8, 9	92%

Adapun analisis dari Tabel 4.2 menyajikan informasi tingkat akurasi specificity algoritma K-Nearest Neighbor. Analisis dilakukan dengan melakukan perhitungan Jumlah benar / jumlah data \* 100%.

Akurasi adalah presentase dari jumlah total prediksi yang benar pada proses klasifikasi (Deng et al, 2016). Hal ini dilakukan berdasarkan table of Confusion tiap-tiap kelas pada Confusion Matrix yang diperoleh pada hasil pelatihan dan pengujian.



**Gambar 5.** Hasil Pengujian Variasi Nilai K Metode K-NN dengan 30 Data Uji

Pada gambar 4.1 diatas terlihat bahwa dari nilai K 1 hingga 10 yang di uji persentase dari hasil analisis metode K-NN lebih tinggi dari hasil analisis metode K-NN. Dan dari nilai K yang telah di ujikan nilai K 2 dan nilai K 9 memiliki persentasi terbesar sehingga keakuratanya juga semakin tepat.

Adapun dari hasil pengujian metode K-Nearest Neighbor dalam klasifikasi data. Adapun pengujian penulis menggunakan variasi nilai K K-Nearest Neighbor 3,4,5,6,7,8,9. Memiliki persentase akurasi yang sangat baik di dibandingkan hanya K-NN. Hasil pengujian menunjukkan metode K-Nearest Neighbor dalam klasifikasi data memiliki akurasi persentase yang baik ketika menggunakan data random. Persentase variasi nilai K K-Nearest Neighbor 3,4,5,6,7,8,9 memiliki persentase 100 % .

## KESIMPULAN

1. Sistem yang dibangun dapat memudahkan dalam proses K-Nearest Neighbor untuk mengetahui kinerja dan peningkatan akurasi dalam klasifikasi citra.





2. Hasil pengujian metode K-Nearest Neighbor Adapun terlihat bahwa dari nilai K 1 hingga 10 yang di uji persentase dari hasil analisis metode K-NN lebih tinggi dari hasil analisis metode K-NN. Dan dari nilai K yang telah di ujikan nilai K 2 dan nilai K 9 memiliki persentasi terbesar sehingga keakuratanya juga semakin tepat. Adapun dari hasil pengujian metode K-Nearest Neighbor dalam klasifikasi data. Adapun pengujian penulis menggunakan variasi nilai K K-Nearest Neighbor 3,4,5,6,7,8,9. Memiliki persentase akurasi yang sangat baik di dibandingkan hanya K-NN. Hasil pengujian menunjukkan metode K-Nearest Neighbor dalam klasifikasi data memiliki akurasi persentase yang baik ketika menggunakan data random. Persentase variasi nilai K K-Nearest Neighbor 3,4,5,6,7,8,9 memiliki persentase 100 % .

#### DAFTAR PUSTAKA

- Boiculese, V.L., Dimitru, G., & Moscalu, M. 2013. Improving Recall of K-Nearest Neighbor Algorithm for Classes of Uneven Size. *The 4th IEEE International Conference on E-Health and Bioengineering-EHB* : 1-4.
- Haviluddin, “Memahami Penggunaan UML (Unified Modelling Language)”, Jurnal Informatika Mulawarman, Vol. 6, No. 1, Februari 2011.
- Jaafar, H., Mukahar, N., & Ramli, D.A. 2016. *Methodology of Nearest Neighbor: Design and Comparison of Biometric Image Database*. IEEE Student Conference on Research and Development (SCORED) : 1-6.
- Pan, D., Zhao, Z., Zhang, L., & Tang, C. 2017. *Recursive Clustering K-Nearest Neighbors Algorithm and the Application in the Classification of Power Quality Disturbance*. IEEE Conference on Energy Internet and Energy System Integration (EI2) : 1-5.
- Priyanto Hidayatullah, 2012, “*Visual Basic .NET Membuat Aplikasi Database dan Program Kreatif*”. Bandung : Informatika.

<https://informatikalogi.com/algorithm-k-nn-k-nearest-neighbor/>

<https://www.ketutrare.com/2018/11/algorithm-k-nearest-neighbor-dan-contoh-soal.html>