

Pengembangan Model Pembelajaran Mesin Prediksi Kematangan Buah Pisang Berdasarkan Citra Digital

Theresia Samantha

Informatika, Fakultas Industri Kreatif Institut Teknologi dan Bisnis Kalbis
Jalan Pulomas Selatan Kav. 22, Jakarta 13210
Email: theresiasamantha32@gmail.com

Abstract: This study aims to develop machine learning to detect banana ripeness based on the skin color. In addition to the color of the skin, also carried out a comparison between the color of the banana with black spots on the banana. The color value used is RGB Color values are taken by using k – means to take the most dominant color value on the banana. After the data is obtained it will be learned using k - nearest neighbor. So as to produce a machine learning to get accuracy for this detection compared with 80% of the trained data and 20% of the tested data. So that the accuracy of the data that is only based on color is compared with the result value is 95.74% with the data using color data and the percentage of the color dominance value is 51.06%.

Keywords: Machine Learning, Computer Vision, KNN, K-Means

Abstrak: Tujuan dari penelitian ini membangun model pembelajaran mesin untuk mendeteksi kematangan buah pisang berdasarkan warna kulitnya. Selain warna kulitnya, juga dilakukan perbandingan antara warna pisangnya dengan bercak kehitaman pada buah pisang tersebut. Nilai warna yang digunakan adalah RGB. Nilai warna diambil dengan cara menggunakan K – Means untuk mengambil nilai warna yang paling dominan pada buah pisang tersebut. Setelah data didapatkan akan dipelajari menggunakan K – Nearest Neighbor. Sehingga menghasilkan pembelajaran mesin untuk mendapatkan akurasi untuk pendeteksian ini dengan perbandingan data yang dilatih sebanyak 80% dan data yang diuji sebanyak 20%. Sehingga dilakukan perbandingan akurasi antara data yang hanya berdasarkan warna saja dengan nilai hasilnya 95.74% dengan data yang menggunakan data warna dan persen nilai dominasi warna dengan nilai hasilnya 51.06%.

Kata kunci: Pembelajaran Mesin, Citra Digital, KNN, K-Means

I. PENDAHULUAN

Buah pisang merupakan salah satu buah tropis yang paling umum ada di Indonesia. Sehingga buah ini juga merupakan salah satu buah yang sering dikonsumsi oleh orang Indonesia. Supaya bisa menikmati buah pisang, maka perlu diketahui apakah kematangan buah pisang itu sudah pas atau belum. Pada umumnya, manusia mengetahui tingkat kematangan buah pisang dengan melihat warna kulit buah pisang

Dalam industri, diperlukannya mempertahankan kualitas dari barang yang dijual. Kualitas buah biasanya dipilih langsung oleh petani yang memanen buah pisang tersebut. Namun, pendapat setiap orang mengenai tingkat kematangan buah pisang tersebut berbeda – beda. Oleh karena itu diperlukannya juga analisis sampai batas mana buah pisang itu bisa dimakan.

Sebelumnya telah ada beberapa penelitian mengenai pendeteksian buah pisang ini. Salah satunya adalah yang dilakukan oleh Mohammad Faizal Ajizi dan timnya telah

melakukan penelitian mendeteksi buah pisang berdasarkan RGB dan beratnya melalui jurnal “Klasifikasi Kematangan Buah Pisang Berbasis Sensor Warna Dan Sensor Load Cell Menggunakan Metode Naive Bayes”. [1] Ada juga yang menggunakan warna lain selain RGB yaitu HIS. Yang terdapat pada jurnal penelitian “Deteksi Kematangan Buah Pisang Berdasarkan Fitur Warna Citra Kulit Pisang Menggunakan Metode Transformasi Ruang Warna HIS” oleh Indarto dan Murinto. [2] Serta beberapa penelitian lainnya. Baik itu dari buah pisang maupun buah yang lainnya.

Dalam penelitian ini, peneliti ingin mendeteksi kematangan buah pisang menggunakan *computer vision* untuk mendeteksi warna kulit pisang tersebut. Dataset berupa gambar yang akan diubah menjadi angka yang merupakan nilai warna buah tersebut. Pengambilan nilai warna pada buah pisang tersebut menggunakan $k - means$. Dengan menggunakan $k - means$ diambil nilai dominan antara warna buah pisang dengan bercak kehitaman pada buah pisang tersebut. Persen warna yang diambil adalah warna buah pisang yang paling dominan pada buah pisang tersebut.

Lalu data tersebut dimasukan ke dalam *testing* dan *training*. Data di *testing* dan *training* dengan menggunakan algoritme *K - Nearest Neighbors* (KNN). Tujuan dari memasukkan data *testing* dan *training* adalah untuk mengetahui akurasi dari klasifikasi kematangan tersebut. Sehingga kita mengetahui hubungan warna buah pisang tersebut dengan warna bercak buah pisang tersebut berhubungan atau tidak dengan klasifikasi kematangan buah pisang tersebut.

II. METODE PENELITIAN

A. Tinjauan Pustaka

1. Computer Vision

Pelihatian komputer (*computer vision*) adalah bidang yang berhubungan dengan upaya untuk membuat komputer dapat menangkap informasi yang berada dalam suatu gambar atau video. [3] *Computer vision* adalah ilmu komputer yang memperoleh, mengolah, dan menganalisis suatu gambar digital. *Computer vision* mengambil inspirasi dari cara manusia

melihat. Yang kemudian mencoba diterapkan ke dalam komputer.

Pada umumnya, manusia mengambil suatu gambar melalui mata mereka. Lalu diproses melalui otak mereka. Didalam otak manusia, sudah terlatih untuk mengenal objek sejak bayi. Sehingga manusia bisa mengenal suatu objek yang dia lihat tersebut. [4]

Untuk komputer, mengambil objek diperlukan kamera untuk menangkap objek tersebut yang berkerja seperti mata manusia. Lalu dari hasil kamera tersebut, dihasilkan berupa gambar atau video. Dari data tersebut diproses melalui komputer yang telah melalui pembelajaran seperti otak manusia.

Biasanya manusia sudah bisa mengenal suatu objek dalam bentuk tiga dimensi. Sehingga kita bisa mengenal bentuk objek tertentu. Serta kita bisa membedakan antara objek dan latar belakangnya. Sehingga di *computer vision* diperlukannya perhitungan matematika untuk mengenal bentuk dan wujud suatu objek. [5]

2. Ruang Warna RGB

RGB merupakan singkatan dari *red*, *green*, dan *blue*, yang memiliki arti merah, hijau, dan biru. Seperti warna dasar pada Hue. Ketiga warna ini juga yang ada disetiap piksel yang mewakili setiap warna pada piksel tersebut. Nilai persetiap warnanya mulai dari 0 sampai 255. Ketika semua warna bernilai 0 atau tingkat terendah, maka warna piksel itu berwarna hitam. Begitu juga sebaliknya, kalau semua warna bernilai 255 atau tingkat tertinggi maka warna piksel tersebut adalah putih. Pada citra digital, urutan asli dalam setiap piksel bukannya RGB namun BGR (*Blue*, *Green*, dan *Red*). Oleh karena itulah, perlu diproses lagi sehingga sesuai dengan urutan yang diperlukan. [3]

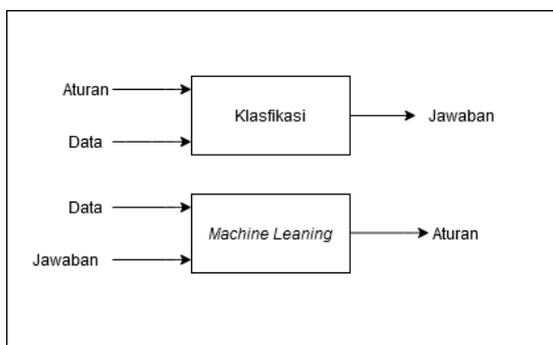
3. Machine Learning

Machine learning atau juga disebut juga sebagai pembelajaran mesin merupakan salah satu bagian dari cakupan artificial intelligence. Kecerdasan buatan atau bisa disebut dengan artificial intelligence adalah mengacu pada bagaimana komputer bisa melakukan pekerjaan yang bisa dilakukan manusia. Sehingga komputer tersebut bisa berpikir sendiri untuk

melakukan tugas yang seperti dilakukan oleh manusia.

Machine learning tidak sama dengan statistik. Karena machine learning memiliki jumlah yang lebih banyak dari pada statistic.[6] Kadang kala banyak orang yang masih salah paham antara machine learning dengan statistik karena sama – sama memprediksi data untuk memprediksi yang akan terjadi di masa depan. Statistic juga yang menjadi fondasi terciptanya machine learning. Karena beberapa bagian dari machine learning mengambil beberapa teori statistik untuk membangun algoritmenya.[7]

Machine learning berbeda dengan klasifikasi biasa. Machine learning memiliki tujuan untuk mempelajari hubungan antara data yang dimasukkan dengan target datanya. Sehingga, membutuhkan data untuk dilatih dan mencobanya pada data yang ada untuk menemukan perbedaan yang membedakan satu hal dengan hal lainnya. Sedangkan klasifikasi biasa sudah mempunyai aturan tertentu untuk membedakan satu hal dengan hal lainnya.[6] Seperti pada Gambar 1.



Gambar 1 Perbedaan Klasifikasi dan *Machine Learning*

4. K - Means

K – means adalah salah satu algoritme untuk pembelajaran unsupervised learning. Algoritme ini akan melakukan pembelajaran dengan melakukan pengelompokan yang berjumlah sesuai dengan jumlah clustering yang telah ditentukan. Pengelompokan dilakukan dengan memberi satu titik di setiap pengelompokan data secara acak. Titik ini disebut centroid.

Centroid akan mencari titik lain yang merupakan data yang dilatih. Titik data yang paling dekat dengan centroid pada kelompok tersebut, maka data tersebut akan dimasukkan ke dalam kelompok tersebut.[8]

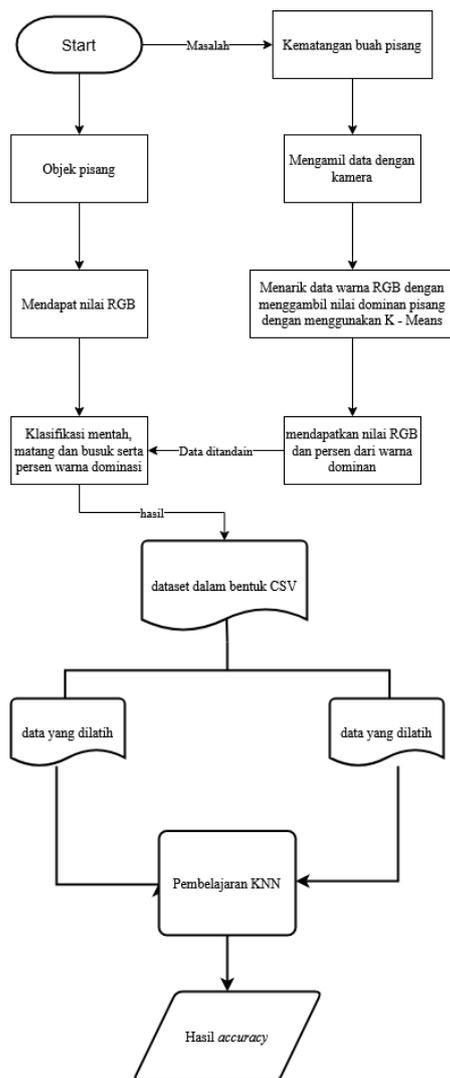
5. K – Nearest Neighbor

K – Nearest Neighbor atau biasanya disingkat KNN merupakan salah satu algoritme untuk pembelajaran supervised learning. *K – Nearest Neighbor* adalah pengelompokan data berdasarkan sifat yang sama dengan kelompok tertentu. KNN memiliki prinsip yang hampir sama dengan k – means.[9] Namun, perbedaannya terletak pada titik pusat.

Patokan pada KNN bukan berdasarkan pada titik pusat, melainkan patokan klasifikasi ada pada data data yang dilatih. Saat data yang ingin dicoba dimasukkan ke dalam KNN, data tersebut akan diukur jaraknya dengan data yang dilatih di sekitarnya. Ketika data yang diuji itu lebih dekat dari data klasifikasi tertentu, maka data yang diuji tersebut masuk ke dalam klasifikasi data yang dekat tersebut.

B. Proses Penelitian

1. Kerangka Pemikiran



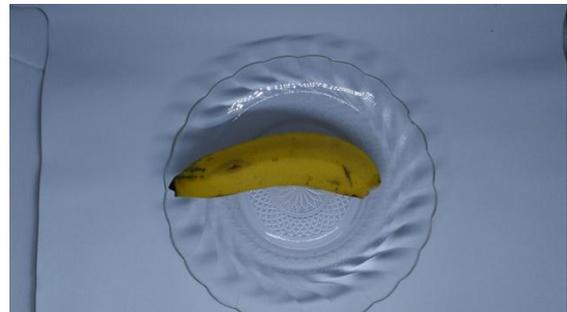
Gambar 2 Kerangka Pemikiran

Pada Gambar 2 menunjukkan diagram kerangka pemikiran yang digunakan pada penelitian ini. Kerangka pemikiran ini menjelaskan tahap demi tahap proses penelitian yang dilakukan pada penelitian ini. Yang dimulai dengan pengambilan data sampai hasil perhitungan akurasi dari penelitian ini.

2. Pengambilan data

Data diambil dengan memfoto buah pisang untuk menjadi citra digital. Gambar buah pisang diambil dalam jangka waktu tertentu. Pengambilan gambar diambil diawali dengan memfoto pisang dari mentah hingga busuk. Lalu, gambar dikelompokkan berdasarkan klasifikasinya. Klasifikasi terdiri dari 3 yaitu mentah, matang, dan busuk.

Kemudian, gambar diproses lagi untuk mendapatkan nilai warnanya. Dimulai dengan memotong gambar sesuai dengan ukuran objeknya.



Gambar 3 awal



Gambar 4 setelah dipotong

Gambar yang telah dipotong sesuai dengan objek kemudian dimasukkan ke dalam pembelajaran *k - means*. Tujuan pembelajaran *k - means* adalah mendapatkan nilai warna dan persen dominan pada buah pisang tersebut.

Pada *k - means*, dilakukan *clustering* untuk mendapat warna dominan tersebut. *Clustering* terdiri dari 3 yaitu warna dominan pisang, warna bercak hitam, dan latar belakang yang pasti berwarna putih. Sehingga persen yang dihasilkan juga terbagi menjadi 3. Oleh karena itu, dilakukan perhitungan untuk mendapatkan nilai persen sesungguhnya dominan pada buah pisang tersebut.

$$\text{Persen objek} = \frac{\text{tertinggi pertama}}{(\text{tertinggi pertama} + \text{tertinggi kedua})} \quad (1)$$

Persen objek didapatkan dengan persen tertinggi pertama dengan total persen tertinggi pertama dan tertinggi kedua. Alasan mengambil persen tertinggi pertama dan tertinggi kedua karena persen tersebut yang mewakili nilai objek tersebut.

Setelah mendapat nilai warna RGB dan persen warna dominan pada buah pisang tersebut, data dimasukkan ke dalam data CSV. Data yang disimpan ke dalam CSV tersebut

sudah dikelompokkan berdasarkan klasifikasinya.

3. Memproses Data

Setelah data disimpan, data kemudian diproses ke dalam pembelajaran *KNN*. Data yang disimpan tersebut kemudian dibagi sebagai data yang dilatih dan data yang diuji. Perbandingan data yang dilatih dengan data yang diuji adalah 80% dan 20%.

Untuk mengukur akurasi pada klasifikasi pada data tersebut digunakannya *Confusion matrix* untuk mengujinya. *Confusion matrix* adalah tabel yang memvisualisasikan hasil dari kinerja klasifikasi.[10] Bentuk *Confusion matrix* seperti Tabel 1.

		Kelas yang di prediksi	
		Positif	Negatif
Kelas sebenarnya	Positif	<i>True Positive (TP)</i>	<i>True Negative (TN)</i>
	Negatif	<i>False Positive (FP)</i>	<i>False Negative (FN)</i>

Tabel 1 Confusion matrix

Untuk menghitung akurasi dari tabel tersebut menggunakan rumus 2.

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

Akurasi dapat didapatkan dengan membagi antara penjumlahan *true positive* dan *true negative* dengan total seluruh data.

III. HASIL DAN PEMBAHASAN

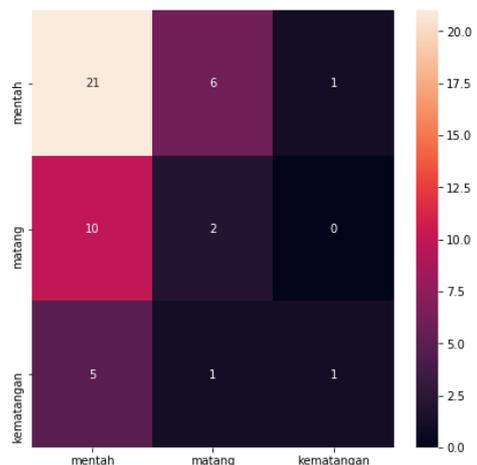
Pada Tabel 2 ini menunjukkan sedikit bagian dari hasil pengambilan nilai warna RGB dan persen warna dominan pada setiap gambar. Diambil data sebanyak 10 dari 231 data. Data telah dikelompokkan sesuai dengan klasifikasinya. Setiap nomor class mewakili setiap klasifikasi. Nomor 0 mewakili kelas mentah, nomor 1 mewakili kelas matang, dan nomor 2 mewakili kelas busuk.

red	green	blue	Persen	class
-----	-------	------	--------	-------

114	128	16	62	1
38	47	160	55	1
103	63	20	70	2
28	30	152	63	1
55	42	150	76	0
47	21	18	83	0
68	63	16	74	1
53	83	13	52	0
106	66	4	52	0
120	33	16	66	0

Tabel 2 hasil menarik data

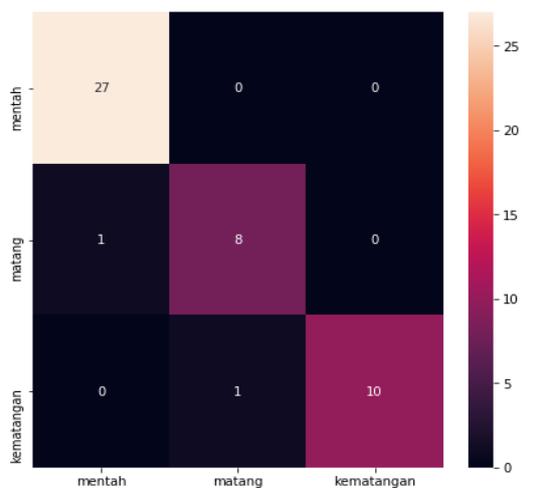
Dilakukan perbandingan akurasi antara data yang menggunakan persen dengan data yang hanya menggunakan nilai warna RGB saja. Pada Gambar 5 menunjukkan hasil *confusion matrix* pada data nilai warna RGB dan persennya.



Gambar 5 Confusion matrix dengan nilai RGB dan persen

$$akurasi = \frac{21 + 2 + 1}{21 + 6 + 1 + 10 + 2 + 0 + 5 + 1 + 1} = 0.5106 * 100\% = 51.06\%$$

Sehingga didapatkan persen berjumlah 51.06%. Lalu, pada Gambar 6 menunjukkan hasil *confusion matrix* pada data yang hanya memiliki nilai warna RGB saja.



Gambar 6 Confusion matrix dengan nilai RGB saja

$$akurasi = \frac{27 + 8 + 10}{21 + 0 + 0 + 1 + 8 + 0 + 0 + 1 + 10} = 0.9574 * 100\% = 95.74\%$$

Sehingga didapatkan persen berjumlah 95.74%. Dari hasil perhitungan, membuktikan bahwa hanya membuktikan nilai RGB saja yang mempengaruhi klasifikasi tingkat kematangan buah pisang.

IV. SIMPULAN

- Penelitian ini menggunakan dua metode yaitu *k-means* dan *KNN*. *K-means* untuk mendapatkan nilai warna dominan pada gambar objek buah pisang tersebut dan *KNN* untuk melakukan pengujian klasifikasi.
- Setelah dimasukkan ke dalam pembelajaran *KNN*, dilakukan pengukuran klasifikasi menggunakan *confusion matrix*.
- Dilakukan perbandingan akurasi antara data yang menggunakan nilai warna RGB dan persen dengan data yang hanya menggunakan nilai warna RGB saja.
- Dari hasil perhitungan akurasi, didapatkan akurasi pada data dengan nilai warna RGB dan persen sekitar 51.06% sedangkan data dengan nilai warna RGB saja sekitar 95.74%. Sehingga membuktikan bahwa hanya membuktikan nilai RGB saja yang mempengaruhi klasifikasi tingkat kematangan buah pisang.

DAFTAR RUJUKAN

- [1] M. F. Ajizi, D. Syaury, M. Hannats, dan H. Ichsan, "Klasifikasi Kematangan Buah Pisang Berbasis Sensor Warna Dan Sensor Load Cell Menggunakan Metode Naive Bayes," vol. 3, no. 3, hal. 2472–2479, 2019.
- [2] Indarto dan Murinto, "Deteksi Kematangan Buah Pisang Berdasarkan Fitur Warna Citra Kulit Pisang Menggunakan Metode Transformasi Ruang Warna HIS (Banana Fruit Detection Based on Banana Skin Image Features Using HSI Color Space Transformation Method)," *J. Ilm. Inform.*, vol. V, no. November, hal. 15–21, 2017.
- [3] A. Kadir, *Langkah Mudah Pemrograman OpenCV & Python*. Jakarta: Elec Media Komputindo, 2019.
- [4] K. Demaagd, A. Oliver, N. Oostendorp, dan K. Scott, *Practical Computer Vision with SimpleCV*. O'Reilly Media, 2012.
- [5] R. Szeliski, *Computer Vision Algorithms and Applications*. 2011.
- [6] F. Chollet, *Deep Learning with Phyton*. 2018.
- [7] I. H. Witten, E. Frank, M. A. Hall, dan C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. Cambridge, MA: Morgan Kaufmann Publishers, 2016.
- [8] G. Hackeling, *Mastering Machine Learning with Scikit-learn: Apply Effective Learning Algorithms to Real-World Problems Using Scikit-learn*. Packt Publishing Ltd., 2014.
- [9] J. W. G. Putra, *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning*, no. March 2018. Tokyo, 2019.
- [10] EMC Education Services, *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. Wiley, 2105.