

KLASIFIKASI BUAH BELIMBING BERDASARKAN CITRA RED-GREEN-BLUE MENGUNAKAN KNN DAN LDA

Retno Nugroho Whidhiasih¹, Nursinta Adi Wahanani², Supriyanto³

¹Program Studi Teknik Komputer Universitas Islam "45" Bekasi, ²Badan Tenaga Nuklir Nasional Tangerang,

³Program Studi Teknik Informatika Universitas Pakuan Bogor.

Email : ¹retno.nw@gmail.com, ²sintaadi@batan.go.id, ³supriyanto_ipb@yahoo.com

ABSTRAK

Tulisan ini dilakukan untuk membandingkan metode klasifikasi K-nearest neighbour (KNN) dan Linear Discriminant Analysis (LDA) dengan variabel R-G dan R-G-B dari citra buah belimbing untuk memprediksi tingkat kemanisan buah belimbing. Pengenalan digunakan untuk mengelompokkan buah belimbing menjadi tiga kelas yaitu kelas manis, sedang dan asam. Pada tahapan pra proses dilakukan reduksi citra dengan menggunakan analisis komponen utama 2 dimensi (2D-PCA). Percobaan dilakukan dengan menggunakan 3 *fold cross validation*. Hasil yang didapatkan dari penelitian ini menyatakan bahwa Klasifikasi ini dapat digunakan untuk mengklasifikasi belimbing kelas asam dengan tepat sedangkan kelas lainnya ketepatannya tidak mencapai 100%. Metode KNN dengan variabel RG menghasilkan akurasi sebesar 80 %, sedang KNN dengan variabel RGB menghasilkan akurasi sebesar 91 %. Teknik LDA linier maupun LDA dengan ukuran jarak mahalalanobis menghasilkan akurasi sebesar 91 %.

Kata kunci : KNN, LDA, RG, RGB, Klasifikasi Belimbing

ABSTRACT

This paper want to compare classification method between k-nearest neighbor (KNN) and linier discriminant analysis (LDA) from starfruit image for sweetness prediction of starfruit. Recognition is used to classify starfruit in sweet, medium and sour class. Classification is done with R-G and R-G-B predictor feature with 3 *fold cross validation*. Image reduction is then conducted in the Pre-process step by using 2D principal componen analysis. Result of this research denotes this classifier can classify sour starfruit class precisely while sweet and medium classes are less precise. Accuracy of KNN method with RG variable is 81%, while KNN, LDA linier and LDA with mahalalanobis distance with RGB variable is 91%.

Keyword: KNN, LDA, RG, RGB, starfruit classification

¹ Penulis untuk korespondensi

1. Pendahuluan

Tumbuhan belimbing manis (*Averrhoa carambola* Linn.), dikenal dengan beberapa nama seperti; starfruit (bahasa Inggris), belimbing amis (Sunda), blimbing legi (Jawa), bainang sulapa (Makasar), dan balireng (Bugis) (Wiryowidagdo S, dan Sitanggang M 2002). Untuk menjamin mutu dan meningkatkan daya saing produk maka buah belimbing digolongkan ke dalam 3 kelas mutu, yaitu kelas super, kelas A dan kelas B (Badan Standarisasi Nasional 2009).

Untuk melakukan pemutuan buah berdasarkan rasa dilakukan dapat dilakukan dengan melakukan pengukuran Total Padatan Terlarut (TPT) di Laboratorium. Kelemahan dari teknik ini adalah buah belimbing harus di belah (desruktif) untuk dikeahui TPT nya. Untuk mengoptimisasi teknik pemutuan maka dikembangkanlah teknik pemutuan buah dengan menggunakan citra buah. Teknik yang dilakukan adalah dengan mengetahui korelasi antara citra RGB dengan TPT yang dikandung oleh buah belimbing.

Penelitian – penelitian pemutuan buah belimbing yang telah dilakukan diantaranya oleh Abdullah et al. (2005), Zaki (2009), Buono A dan Irmansyah (2009) dan Irmansyah (2009).

Penelitian ini bertujuan untuk membandingkan metode *k-nearest neighbor* (KNN) dan *Linear Discriminan Analysis* (LDA) dengan menggunakan citra RGB dengan pra-proses dengan menggunakan analisis komponen utama (PCA).

Penelitian ini disajikan dengan susunan sebagai berikut : Bagian 2 mengenai teknik reduksi dimensi dengan Analisis Komponen Utama (PCA), Bagian 3 membahas teknik pengenalan pola yang digunakan yaitu KNN dan LDA. Deskripsi mengenai data buah belimbing yang digunakan dalam percobaan ini disajikan pada bagian 4. Bagian 5 akan menyajikan rancangan percobaan, data yang digunakan serta hasilnya. Selantunya kesimpulan dan saran disajikan pada bagian akhir.

2. Analisis Komponen Utama (PCA)

Prosedur PCA pada dasarnya adalah bertujuan untuk menyederhanakan variabel yang diamati dengan cara menyusutkan (mereduksi) dimensinya (Abdi dan Williams 2010). Hal ini dilakukan dengan cara menghilangkan korelasi diantara variabel bebas melalui transformasi variabel bebas asal ke variabel baru yang tidak berkorelasi sama sekali atau yang biasa disebut dengan principal component.

Setelah beberapa komponen hasil PCA yang bebas multikolinearitas diperoleh, maka komponen-komponen tersebut menjadi variabel bebas baru yang akan diregresikan atau dianalisa pengaruhnya terhadap variabel tak bebas (Y) dengan menggunakan analisis regresi.

Maksud dari PCA adalah untuk menangkap variasi total di dalam kumpulan karakter yang dilatihkan, dan untuk merepresentasikan variasi ini dengan variabel yang lebih sedikit. Suatu image yang

direpresentasikan dengan variabel yang sedikit akan lebih mudah untuk ditangani dan dimengerti daripada jika direpresentasikan dengan raw pixel yang banyak dari image tersebut.

Apabila didefinisikan sebuah objek $u = \{u_1, u_2, u_3, \dots, u_n\}$ sebagai vektor pada n dimensi. Objek u dapat berupa suatu gambar dan mempunyai komponen $u_1, u_2, u_3, \dots, u_n$, dimana $u_1, u_2, u_3, \dots, u_n$ adalah nilai pixel dari gambar tersebut. Dengan kondisi ini maka n dapat diartikan sebagai jumlah pixel (=panjang x lebar) yang terdapat dalam gambar. Kemudian, apabila objek tersebut ditambah dengan objek-objek yang lain hingga menjadi sekumpulan atau sekelompok objek maka : $u_i = \{u_{i1}, u_{i2}, \dots, u_{in}\}$, dimana $i = 1, \dots, m$ dan $m < n$.

3. Pengenalan Objek

Teknik pengenalan obyek yang digunakan dalam penelitian ini adalah teknik *k-nearest neighbor* (KNN) dengan menggunakan ukuran jarak Euclid dan *Linear Discriminant Analysis* (LDA).

3.1. K Nearest Neighbor

Algoritma *k-nearest neighbor* (Pencarian tetangga terdekat) merupakan teknik klasifikasi yang sangat populer yang diperkenalkan oleh Fix dan Hodges (1951), yang telah terbukti menjadi algoritma sederhana yang baik. KNN merupakan salah satu metode yang digunakan dalam pengklasifikasian dengan menggunakan algoritma *supervised* (Chan *et al.* 2010).

Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan

jarak suatu obyek yang akan diklasifikasikan terhadap data contoh. *Classifier* hanya menggunakan fungsi jarak dari data baru ke data training.

Prinsip kerja *K-Nearest Neighbor* (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (*neighbor*) terdekatnya dalam data pelatihan. Data pelatihan diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pelatihan. Sebuah titik pada ruang ini ditandai kelas c , jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak Euclidean dengan rumus sebagai berikut :

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

Dengan:

x_1 = sampel data, x_2 = data uji, i = variabel data, $dist$ = jarak, p = dimensi data

Pada fase pembelajaran, algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk data test (yang klasifikasinya tidak diketahui). Jarak dari vektor yang baru ini terhadap seluruh vektor data pembelajaran dihitung, dan sejumlah k buah yang paling dekat diambil. Titik yang baru klasifikasinya

diprediksikan termasuk pada klasifikasi terbanyak dari titik-titik tersebut.

Nilai k yang terbaik untuk algoritma ini tergantung pada data. Umumnya, nilai k yang tinggi akan mengurangi efek *noise* pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Nilai k yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *cross-validation*. Kasus khusus di mana klasifikasi diprediksikan berdasarkan data pembelajaran yang paling dekat (dengan kata lain, $k = 1$) disebut algoritma *nearest neighbor*.

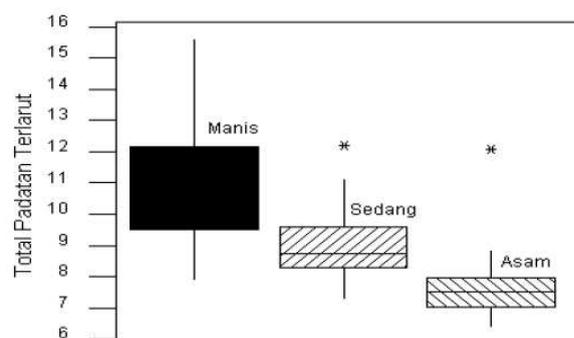
Ketepatan algoritma KNN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. Ketika jumlah data mendekati tak hingga, algoritma ini menjamin *error rate* yang tidak lebih dari dua kali *Bayes error rate* (tingkat minimum pada distribusi data tertentu).

3.2. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) atau analisis diskriminan merupakan klasifier dimana dari data yang ada terdapat beberapa data yang sudah diketahui kelasnya atau labelnya. Data yang sudah diketahui labelnya digunakan untuk menemukan fungsi diskriminan. Untuk data yang belum diketahui kelasnya/labelnya, kita menggunakan fungsi diskriminan yang telah ditemukan.

4. Data Percobaan

Data yang digunakan pada penelitian ini adalah data berformat .bmp hasil pengolahan citra buah belimbing manis yang dilakukan oleh Irmansyah (2008), yang terdiri dari 4 tahap pemanenan (usia 40, 50, 60, dan 70 hari), masing-masing 75 buah. Pengelompokan data didasarkan pada Total Padatan Terlarut (TPT) yang diukur melalui pengamatan Laboratorium. Dalam hal ini makin tinggi nilai TPT maka rasa buah makin manis. Gambar 1 menyajikan perbandingan *boxplot* nilai TPT antara buah manis, sedang dan asam dari beberapa buah yang diuji cita rasa kemanisannya (Buono A dan Irmansyah 2009).



Gambar 1. Boxplot Tiga Cita Rasa Manis Belimbing

Pengategorian kemanisan buah sesuai nilai TPT [5] menggunakan aturan berikut :

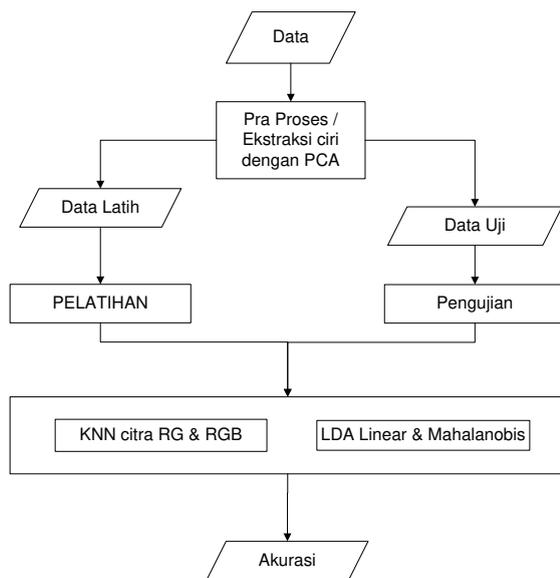
- Manis : $TPT > 7.6$
- Sedang : $5.0 < TPT < 7.6$
- Asam : $TPT < 5.0$

Pada penelitian ini pengenalan pola menggunakan data R-G-B seperti pada penelitian Abdullah *et al.* (2005), Irmansyah (2008), Buono A dan Irmansyah (2009) dan Fathurrahman Z (2009) dari citra yang diperoleh dari pemotretan buah. Fitur yang digunakan pada percobaan ini adalah R-G pada

KNN, R-G-B pada knn, LDA-linier dan LDA-mahalanobis.

5. Rancangan Percobaan dan Hasil

Percobaan ini dilakukan dengan mengikuti alur percobaan yang diperlihatkan pada Gambar 1. Percobaan dilakukan dengan menggunakan 3 fold cross validation seperti yang dilakukan oleh Fathurrahman Z (2009). Mula-mula data dibagi ke dalam 3 bagian, bagian 1 dan 2 digunakan sebagai data pelatihan dan bagian 3 digunakan sebagai data uji. Selanjutnya data 2 dan 3 sebagai data training dan bagian 1 sebagai data uji. Langkah terakhir data 1 dan 3 sebagai data training dan bagian 2 sebagai data uji.



Gambar 2. Alur Proses Percobaan

Data yang digunakan adalah data yang telah direduksi dengan Analisis Komponen Utama (PCA) (Chan L *et al.* 2010). Pada penggunaan metode KNN (Sreemathy J *et al.* 2012) data yang diujikan dibagi ke dalam dua jenis data yaitu data citra R-G dan data citra

RGB. Pada penggunaan metode LDA (Chan L *et al.* 2010) digunakan data R-G-B yang diklasifikasikan dengan LDA Linear dan LDA dengan Jarak Mahalanobis. Selanjutnya data dibagi dua untuk dijadikan data *training* dan data uji.

Tabel 1. Akurasi Hasil Percobaan

Metode	Variabel	Akurasi
k-NN	RG	80.80%
k-NN	RGB	91%
LDA Linier	RGB	91%
LDA mahalanobis	RGB	91%

Tabel 1 menunjukkan akurasi dari hasil percobaan yang dilakukan. Akurasi dari metode klasifikasi dengan menggunakan KNN dengan data citra R-G adalah sebesar 80,80%, akurasi lebih baik pada klasifikasi yang dilakukan dengan citra R-G-B yaitu sebesar 91%. Sementara itu akurasi yang dihasilkan dari teknik klasifikasi dengan menggunakan LDA linier dan LDA – Mahalanobis adalah sebesar 91 %.

Setelah melakukan analisis pada R-G menggunakan metode knn didapatkan akurasi sebesar total 80.80%, akurasi meningkat menjadi 91% setelah dilakukan analisis R-G-B dengan metode yang sama. Analisis R-G-B dibandingkan juga dengan metode LDA-linier dan LDA-mahalanobis, hasil yang didapatkan adalah sama yaitu 91%.

Akurasi yang dihasilkan untuk belimbing asam pada metode k-nn dan LDA adalah sebesar 100% baik dengan R-G maupun R-G-B. Hal ini dapat diartikan bahwa untuk klasifikasi belimbing asam tidak terdapat

kesalahan. Sedangkan untuk manis dan sedang masih terjadi beberapa kesalahan.

Jika dikaitkan dengan penelitian sebelumnya yaitu klasifikasi 1D-PCA dan 2D-PCA yang menyatakan bahwa citra RGB dengan tingkat kemanisan (kandungan TPT) (Buono A dan Irmansyah 2009) memberikan hasil yang sepadan. Hal ini menunjukkan bahwa teknik klasifikasi KNN dan LDA dapat digunakan sebagai alternatif teknik klasifikasi buah belimbing.

6. Kesimpulan Dan Saran

Dari pembahasan yang telah dilakukan dapat diambil beberapa kesimpulan sebagai berikut :

1. Pengelompokan buah belimbing berdasar RGB yang sepadan hasilnya dengan kandungan TPT dapat dikelompokkan menjadi dua kelas, yaitu kelas asam dan kelas tidak asam.
2. Klasifikasi buah belimbing ke dalam tingkat manis, sedang, asam dengan teknik KNN 2 variabel R-G menghasilkan akurasi 80 %. KNN dengan 3 variabel R-G-B menghasilkan akurasi 91 %. Dengan teknik LDA linier maupun Mahalanobis menghasilkan akurasi 91 %
3. Teknik klasifikasi menggunakan KNN dan LDA kurang mampu membedakan antar kelas sedang dan kelas manis.
4. Sebagian besar kesalahan klasifikasi adalah tidak mampu membedakan kelas sedang dan manis

Untuk penelitian selanjutnya dapat dilakukan pemilihan komponen R-G-B yang

tepat agar klasifikasi dengan k-nn dan LDA dapat dilakukan dengan baik dan akurat.

7. Ucapan Terima Kasih

Penulis mengucapkan terimakasih kepada Dr. Agus Buono M.Si, M.Kom, dosen Departemen Ilmu Komputer IPB, atas penggunaan data citra belimbing pada penelitian ini.

8. Referensi

- Abdi H dan Williams LJ. 2010. *Principal Component Analysis. Wiley Interdisciplinary Reviews: Computational Statistics*. 2(4): 433-459.
- Abdullah MZ, Muhammad-Saleh J, Syahir F, dan Azemi M. 2005. Automated inspection system for colour and shape grading of starfruit (averhoa carambola L) using machine vision system. *Transaction of the Institute of measurement and control*. 27(2): 65-87.
- Badan Standarisasi Nasional. 2009. SNI Buah Belimbing 4491:2009. Jakarta (ID): Badan Standarisasi Nasional.
- Buono A dan Irmansyah. 2009. Pengenalan kadar total padat terlarut pada buah belimbing manis berdasar citra RGB dengan analisis komponen utama sebagai ekstraksi ciri dan jarak euclidean sebagai pengenalan pola. *Jurnal Ilmu Komputer dan Informasi*. 2(1).
- Chan L, Salleh S dan Ting C. 2010. Face Biometrics Based on Principal Component Analysis and Linear Discriminant Analysis. *Journal of Computer Science*. 6 (7) : 639-699.

- Fathurrahman Z. 2009. Pengembangan probabilistic neural networks untuk penentuan kematangan belimbing manis [skripsi]. Bogor (ID): Institut Pertanian Bogor.
- Irmansyah. 2008. Evaluasi mutu belimbing dengan pengolahan citra dan logika fuzzy [disertasi]. Bogor (ID): Institut Pertanian Bogor.
- Sreemathy J *et al.* 2012. An efficient text classification using KNN and Naïve Bayesian. *IJCSE*. 4(3): 392-396.
- Wirjowidagdo S, dan Sitanggang M. 2002. Tanaman Obat untuk Penyakit Jantung, Darah Tinggi, dan Kolesterol. Jakarta (ID): AgroMedia Pustaka