

Akurasi Klasifikasi Citra Digital *Scenes* RGB Menggunakan Model K-Nearest Neighbor dan Naive Bayes

1st Elza Fitriana Saraswita
Jurusan Teknik Informatika
Universitas Sriwijaya
Palembang, Sumatera Selatan
Elzfs.9994@gmail.com

2nd Sukemi
Jurusan Teknik Informatika
Universitas Sriwijaya
Palembang, Sumatera Selatan
sukemi@ilkom.unsri.ac.id

Abstrak—Gambar citra digital merupakan larik (array) yang berisi nilai kompleks dengan bit tertentu yang dapat dihitung secara matematis. Pada penelitian ini ada dua metode yang digunakan untuk membandingkan hasil akurasi klasifikasi citra digital RGB yaitu dengan metode Naive Bayes dan K-Nearest Neighbor (KNN). Metode klasifikasi Naive Bayes berdasarkan perhitungan matematika probabilitas yang sederhana dan metode klasifikasi K-Nearest Neighbor (KNN) berdasarkan pada perhitungan kedekatan atau K. Kedua metode tersebut diberikan data set digital citra *scenes* RGB yang sama untuk proses pengelompokan dan pengklasifikasian selanjutnya data akan dilatih dan diuji untuk mendapatkan hasil akurasi. Berdasarkan klasifikasi data set citra digital RGB menggunakan metode K-Nearest Neighbor memiliki tingkat akurasi lebih besar dari metode Naive Bayes. Metode pemrosesan data citra digital menghasilkan Naive Bayes dengan akurasi yang diperoleh adalah sebesar 66% dan metode K-Nearest Neighbor yang diperoleh adalah dengan akurasi 76%. Dari analisis ini dapat disimpulkan bahwa metode K-Nearest Neighbor bekerja lebih baik dibandingkan dengan Naive Bayes untuk data citra digital RGB.

Katakunci—Citra Digital, Klasifikasi, KNN, Naive Bayes, RGB

Abstract—Digital image images are arrays containing complex values with certain bits could be calculated mathematically. In this study there are two methods used to compare the results of the testing RGB digital image classification with the Naive Bayes and K-Nearest Neighbor (KNN) methods. The Naive Bayes classification method of simple mathematical calculations and the K-Nearest Neighbor (KNN) classification methods are based on proximity or K calculations. Both of method is given the same set of RGB digital image scenes data for the process of grouping and classifying the data which will then be drilled and published for get the results of realization. Based on the classification of RGB digital image data sets using the K-Nearest Neighbor method has a greater degree of accuracy than the Naive Bayes method. Digital image data retrieval method produces Naive Bayes with an accuracy of 66% and K-Nearest Neighbor method obtained with an accuracy of 76%. In this analysis it can be denied that the K-Nearest Neighbor method works better than Naive Bayes for RGB digital image data.

Keywords—Digital Image, Classification, KNN, Naive Bayes, RGB.

I. PENDAHULUAN

Teknologi saat ini sangat memungkinkan untuk melakukan klasifikasi citra digital pemandangan (*scenes*) RGB. Secara umum tahapan dalam proses klasifikasi citra digital yaitu akuisisi citra, pra pengolahan citra, ekstraksi ciri/ fitur, pelatihan, pengujian

dan pengukuran akurasi. Tahapan mengekstrak informasi dalam citra digital sangat mempengaruhi untuk mengenali objek yang ada dalam citra tersebut [1]. Naive bayes memiliki kesamaan dengan metode K-Nearest Neighbor (KNN) dalam mengolah jumlah data yang besar akan tetapi, Naive Bayes mengklasifikasikan berdasarkan promentalitas dan teorema bayesian mengasumsikan tiap-tiap variabel bersifat bebas sedangkan metode K-Nearest Neighbor (KNN) merupakan metode klasifikasi yang menentukan label (class) dari sebuah objek baru berdasarkan mayoritas class dari jarak terdekat 'K' dalam kelompok data latih. Klasifikasi gambar pemandangan (*scenes*) diolah dan diklasifikasikan secara kategori. Tidak seperti klasifikasi objek yang berfokus pada pengelompokan objek-objek yang menonjol di latar depan, klasifikasi pemandangan (*scenes*) menggunakan tata letak objek pemandangan dalam pemandangan (*scenes*), selain konteks lingkungan (*ambient*), untuk klasifikasi. Perhitungan komputasi ini menggunakan pengenalan objek pemandangan (*scenes*) dan memperlihatkan dasar dari representasi pemandangan (*scenes*) yang akan di analisis. Dalam data set citra digital pemandangan (*scenes*) yang di ambil untuk penelitian ini memperkirakan struktur atau "bentuk pemandangan" menggunakan beberapa dimensi persepsi yang secara khusus didedikasikan untuk menggambarkan sifat spasial pemandangan.

Pada data visual lebih sulit untuk mengklasifikasikan gambar ketika direduksi menjadi resolusi rendah. Sementara penulis menggunakan seluruh dataset yang dilakukan pada penelitian [2]. Kemudian diputuskan untuk mengambil resolusi 256x256. Penelitian – penelitian pemutuan klasifikasi data citra digital sebelumnya yang telah dilakukan pada penelitian [2], [1], dan [3].

II. METODOLOGI

A. Data Persiapan

Data set gambar yang digunakan dalam penelitian ini adalah corel dataset yang sebelumnya telah dibuat dan dikelola oleh [2] dengan total 8000 gambar pemandangan (*scenes*) yang terdiri dari data gambar digital RGB pemandangan (*scenes*) terdiri dari sekitar 4000 pemandangan alam (misalnya, pesisir, pantai, lautan, pulau, lapangan, gurun, padang rumput,

lembah, danau, sungai, gunung, ngarai, gua, hutan, air terjun, taman, dll.), dan sekitar 3500 lingkungan perkotaan (mis. gedung pencakar langit, pusat kota, area komersial, jalan, jalan raya, rumah, bangunan, pusat pejalan kaki, tempat, parkir, dll.). Sisa gambar sebanyak 600 disesuaikan dengan pemandangan ambigu dalam hal tingkat kealamian (mis., pemandangan pertanian, desa di pegunungan, pemandangan kota dan pemandangan udara, dll.) [2].

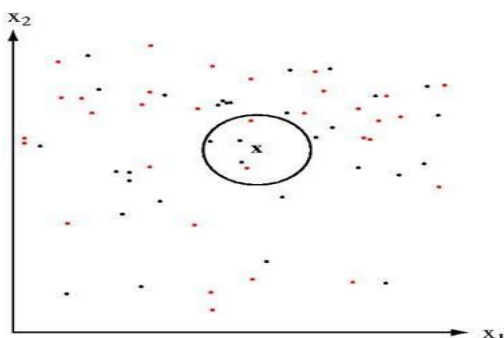
Data yang penulis ambil untuk penelitian ini dari 8000 gambar sebanyak 948 gambar dalam 3 kategori, yaitu pesisir (*coast*) sebanyak 360 gambar, Hutan (*forest*) sebanyak 328 gambar, Jalan raya (*Highway*) sebanyak 260 gambar. Seluruh gambar pada corel dataset tersebut berjenis *truecolor* yang artinya setiap *pixel* gambar memiliki 3 komponen warna yaitu merah (*red*), hijau (*green*), dan biru (*blue*). Gambar - gambar citra digital ini yang didapatkan untuk penelitian ini memiliki resolusi yang cukup besar sehingga dapat meningkatkan kompleksitas perhitungan. Gambar berukuran 256×256 piksel, dalam 256 tingkat abu-abu yang berasal dari perpustakaan foto saham Corel, gambar yang diambil dari kamera digital dan gambar yang diunduh dari web.

B. Scenes

Citra digital pemandangan (*scenes*) dibentuk berdasarkan pertimbangan jarak absolut antara pengamat dan zona terfiksasi. Gambar "objek" direpresentasikan ketika pandangan berada di bawah 1 hingga 2 meter di sekitar pengamat dan suatu pemandangan (*scenes*) dimulai ketika ada ruang yang lebih besar antara pengamat dan titik tetap, biasanya setelah 5 meter [2]. Dengan demikian, sehingga dapat ditetapkan gambar digital citra RGB tersebut sebuah pemandangan (*scenes*) atau gambar *outdoor* Gambar 1.



Gambar 1. Sampel gambar berlabel dari database scenes



Gambar 3. Gambaran umum klasifikasi

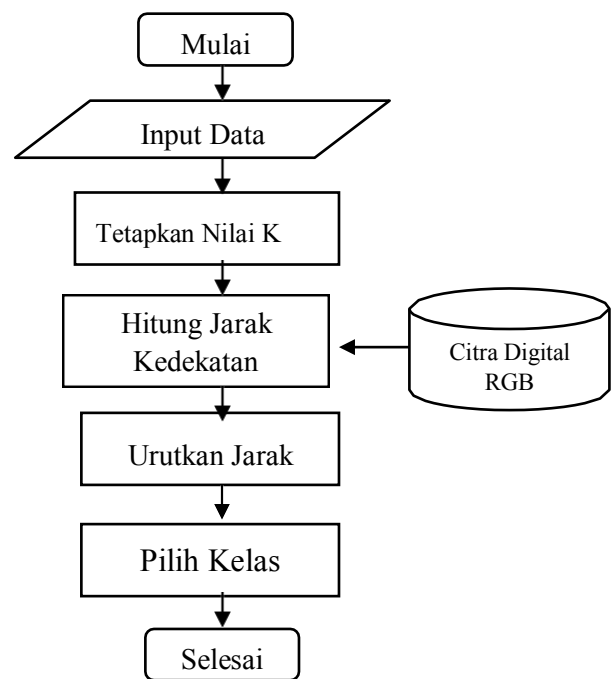
C. Klasifikasi

Klasifikasi adalah suatu proses pengkategorian yang dilakukan terhadap sekumpulan data Gambar 3.

Klasifikasi diawali dengan membagi koleksi data menjadi dua bagian utama yaitu data latih dan data uji. Dari data latih di hitung dengan menggunakan metode tertentu kemudian, diperoleh model klasifikasi yang nantinya akan digunakan untuk penentuan kelas terhadap data uji.

D. K- Nearest Neighbor (KNN)

Pengklasifikasian menggunakan fungsi jarak dari data baru ke data training (kedekatan). Prinsip kerja K-Nearest Neighbor (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (*neighbor*) terdekatnya dalam data pelatihan [4]. 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. Algoritma di perlihatkan pada Gambar 2.



Gambar 2. Algoritma K-Nearest Neighbor

Keterangan algoritma K-Nearest Neighbor:

1. Tentukan parameter K
2. Hitung jarak antara data yang akan dievaluasi dengan semua data latih.
3. Urutkan jarak data
4. Tentukan jarak terdekat sampai urutan K
5. Pasangkan kelas yang bersesuaian
6. Cari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi.

Rumus K-NN ditunjukkan pada persamaan (1)

$$d_1 = \sqrt{\sum_{i=1}^p (X_{2i} - X_{1i})^2} \quad \dots(1)$$

Keterangan:

X_1 = Sampel Data

X_2 = Data Uji / Testing

i = Variabel Data

d = Jarak

p = Dimensi Data

Klasifikasi menggunakan voting terbanyak diantara klasifikasi dari k obyek. Algoritma KNN menggunakan klasifikasi ketetangaan sebagai nilai prediksi dari query instance yang baru. Algoritma metode KNN sangatlah sederhana, bekerja berdasarkan jarak terpendek dari query instance ke training sample untuk menentukan KNN-nya.

E. Naive Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [5]. Metode menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel. Rumus umum teorema bayes ditunjukkan pada persamaan (2).

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad \dots(2)$$

Keterangan:

X :Data dengan class yang belum diketahui

H :Hipotesis data merupakan suatu class spesifik

$P(H|X)$:Probabilitas hipotesis H berdasar kondisi X

$P(H)$:Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$:Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$:Probabilitas X

F. Confusion matrix

Metode ini menggunakan tabel matriks yang digunakan untuk membandingkan jumlah TP terhadap Annual Research Seminar (ARS) 2019
Fakultas Ilmu Komputer UNSRI

$$Precision = \frac{TP}{TP+FP} \quad \dots(3)$$

$$Recall = \frac{TP}{TP+FN} \quad \dots(4)$$

$$F1 = \left(\frac{2}{recall^{-1} + precision^{-1}} \right) \quad \dots(5)$$

Sebuah tabel klasifikasi binary untuk menentukan data yang digunakan bersifat *True Positives (TP)*, *False Positives (FP)*, *False Negatives (FN)*, dan *True Negatives (TN)*. Penentuan dilihat dari TABEL 1.

TABEL 1. TABEL KLASIFIKASI BINARY

No	Klasifikasi yang benar	Diklasifikasikan sebagai	
		+	-
1	+	True Positives (TP)	False Negatives (FN)
2	-	False Positives (FP)	True Negatives (TN)

Keterangan:

TP = Jumlah true positives

TN = Jumlah true negatives

FN = Jumlah false negatives

FP = Jumlah false positives

P = Jumlah record positif

N = Jumlah tupel negatif

III. PENGUJIAN DATA

jumlah record yang positif dengan jumlah TN terhadap jumlah record yang negatif [6]. Untuk menghitung *precision*, *Recall*, dan *F1-score* digunakan persamaan (3), (4), dan (5) di bawah ini:

Pengujian data dilakukan setelah mendapatkan data citra digital RGB yang telah dikelompokkan perkategori. Kemudian, tahap selanjutnya adalah proses pengolahan data untuk dilatih dan diuji agar menghasilkan akurasi menggunakan bahasa python dalam program jupyter notebook. Alur data yang diujikan dalam Gambar 4.

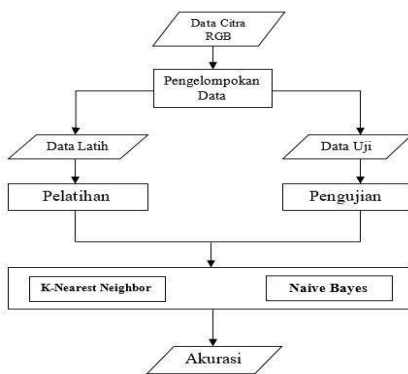
A. Pelatihan

Pada model pelatihan terdapat suatu kumpulan data yang memiliki kelas yang telah ditetapkan (data latih). Data pemandangan (*scenes*) diambil sebanyak 75 persen perkategori. Data citra digital kategori pesisir (*coast*) sebanyak 270 gambar, kategori hutan (*forest*) sebanyak 246 gambar, dan kategori jalan raya (*Highway*)

sebanyak 195 gambar. Kumpulan data latih ini digunakan untuk membentuk model klasifikasi.

B. Pengujian

Pengujian menggunakan model klasifikasi Naive Bayes dan K-Nearest Neighbor (KNN) yang sudah terbentuk dan diujikan dengan data yang lainnya (data uji). Dilakukan untuk mendapatkan hasil *precision*, *recall*, dan *f1-score* hingga, mengetahui akurasi dari tiap model klasifikasi tersebut. Model pengujian data pemandangan (*scenes*) diambil sebanyak 25 persen perkategori. Data citra digital kategori pesisir (*coast*) sebanyak 90 gambar, kategori hutan (*forest*) sebanyak 82 gambar, dan kategori jalan raya (*Highway*) sebanyak 65 gambar.



Gambar 4. Alur pengujian data

A. Evaluasi

Evaluasi merupakan fase lanjutan terhadap data latih dan data uji. Evaluasi dilakukan secara mendalam dengan tujuan agar hasil pada tahap pemodelan sesuai dengan sasaran yang ingin dicapai. Hasil perhitungan pengujian metode K-Nearest Neighbor (KNN) dan Naive Bayes dengan model *precision*, *recall*, dan *f1-score* dapat di lihat pada TABEL 2 dan TABEL 3.

Scenes	Precision	Recall	F1-Score
Coast	0.75	0.70	0.73
Forest	0.81	0.76	0.79
Highway	0.74	0.86	0.79

TABEL 2 TABEL HASIL UJI MODEL METODE KNN

Scenes	Precision	Recall	F1-Score
Coast	0.60	0.73	0.46
Forest	0.81	0.74	0.73
Highway	0.50	0.78	0.61

TABEL 3 TABEL HASIL UJI MODEL METODE NAIVE BAYES

I. HASIL

Klasifikasi citra digital pemandangan (*scenes*) RGB menggunakan metode Naive Bayes dan metode dan K-Nearest Neighbor (KNN) dengan melalui semua tahapan dipastikan tidak ada bagian – bagian penting yang terlewatkan, dihasilkan dari perhitungan akurasi persamaan (6) .

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad \dots(6)$$

Hasil menyatakan bahwa nilai akurasi terhadap klasifikasi metode K-Nearest Neighbor (KNN) sebesar 76% dan metode Naive Bayes sebesar 60%. Persentase dalam akurasi terbilang kecil yang disebabkan oleh kurang kompleksitas data yang mengakibatkan model dapat memprediksi tidak cukup akurat.

II. KESIMPULAN

Penelitian ini ingin membuktikan berapa presentase keberhasilan pengenalan citra digital pemandangan (*scenes*) RGB dan membandingkan tingkat presentase keberhasilan yang lebih baik antara metode Naive Bayes dan metode K-Nearest Neighbor (KNN). Untuk menguji tingkat keberhasilan, penulis menggunakan model perhitungan *Precision*, *Recall*, *F1-Score*, dan Akurasi.

Berdasarkan hasil uji coba yang telah dilakukan pada penelitian ini, terbukti bahwa metode K-Nearest Neighbor dapat digunakan untuk mengklasifikasi citra digital pemandangan (*scenes*) RGB lebih baik dari metode Naive Bayes. Akan tetapi tingkat akurasi masing kurang memuaskan, yaitu 76% dan 66%. Maka, untuk penelitian berikutnya metode dapat dikembangkan lagi. Dengan begitu dapat memaksimalkan proses prapengolahan dan ekstraksi ciri citra.

REFERENCES

- [1] C. Science and Y. T. Pratama, "Optimalisasi Image Analisis Noise Citra Menggunakan Algoritma Gaussian Filter," vol. 4, no. 1, pp. 978–979, 2018.
- [2] A. Oliva, W. Hospital, and L. Ave, "Modeling the Shape of the Scene : A Holistic Representation of the Spatial Envelope *," vol. 42, no. 3, pp. 145–175, 2001.
- [3] A. M. Hasan, F. Meziane, R. Aspin, and H. A. Jalab, "SS symmetry Segmentation of Brain Tumors in MRI Images Using Three-Dimensional Active Contour without Edge," no. November, 2016.
- [4] M. E. I. Lestari, "PENERAPAN ALGORITMA KLASIFIKASI NEAREST NEIGHBOR (K-NN) UNTUK MENDETEKSI PENYAKIT JANTUNG," vol. 7, no. September 2010, pp. 366–371, 2014.
- [5] A. Ciputra, "DENGAN ALGORITMA NAIVE BAYES DAN EKSTRAKSI FITUR CITRA DIGITAL," vol. 9, no. 1, pp. 465–472, 2018.
- [6] E. Ronando, M. I. Irawan, and E. Apriliani, "(SVM) – Neuro Fuzzy For Fast Data Classification," no. August, 2015.