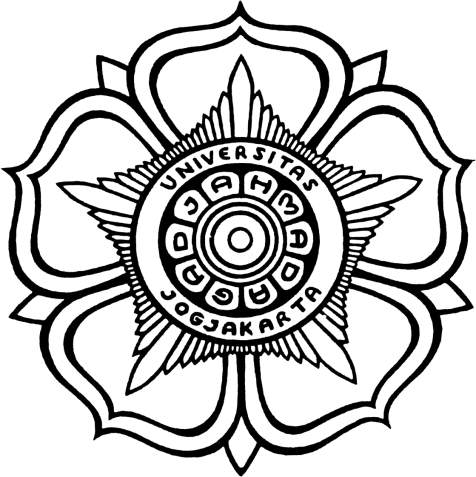
**PENGENALAN POLA**

**KLASIFIKASI DATA GAMBAR ANJING DAN KUCING DENGAN**

**K-NEAREST NEIGHBOR**

****

**ANGGA MAULANA PURBA**

**16/403663/PPA/05180**

**PROGRAM STUDI S2 ILMU KOMPUTER**

**DEPARTEMEN ILMU KOMPUTER DAN ELEKTRONIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS GADJAH MADA**

**YOGYAKARTA**

**2017**

# PENDAHULUAN

Pemrosesan gambar merupakan salah satu fokus ilmu dari pengenalan pola. Pemrosesan gambar secara spesifik dalam sistem ini meliputi tahap preprocessing, ektraksi fitur dan kemudian klasifikasi tanpa melalui segmentasi. Pada tahap preprocessing dilakukan manipulasi data gambar agar data gambar mempunyai dimensi ukuran yang sama. Termasuk di dalam tahap preprocessing ada tahap grayscaling agar data gambar hanya mempunyai intensitas untuk memperkecil kompleksitas pada komputasi. Kemudian tahap selanjutnya dilakukan thresholding terhadap wana dengan intensitas lebih dari nilai tertentu agar data intensitas background tidak disertakan dalam komputasi. Tahap preprocessing ini dilakukan oleh library dari python yang bernama scikit-image, baik untuk grayscaling dan normalisasi.

Fitur yang dihasilkan dari tahap sebelumnya adalah intensitas pada masing-masing piksel. Setiap data gambar akan menghasilkan matriks berukuran 100x100 yang merepresentasikan intensitas (di dalam library direpresentasikan antara 0-1) untuk setiap piksel pada koordinat tertentu. Proses ekstraksi fitur intensitas dilakukan untuk semua gambar, baik itu data uji maupun data training untuk kemudian dibandingkan jarak kemiripannya.

Hasil dari perhitungan matriks gambar yang berukuran 100x100 tersebut yang akan dijadikan sumber dari proses klasifikasi. Algoritma yang digunakan adalah k-Nearest Neighbor dengan nilai k dilakukan eksperimen untuk sistem ini. Algoritma ini bekerja dengan mencari sejumlah k tertangga terdekat dari data uji. Tetangga mayoritas dengan kelas tertentu akan dijadikan kelas dari data uji tersebut. Ketetanggaan dalam algoritma ini didasarkan dari jarak tertentu yang dihitung dengan jumlah selisih absolut dari 2 data atau biasa disebut L1 distance.

# METODE

## PREPROCESSING DAN EKTRAKSI FITUR

### Grayscaling

Citra Grayscale, suatu citra dikatakan sebagai citra grayscale apabila sebuah citra tidak memiliki warna RGB atau dapat dikatakan sebuah citra yang memiliki nilai dari putih yang memiliki intensitas paling besar sampai hitam yang memiliki intensitas paling rendah. Citra grayscale terdiri dari x dan y dalam kordinat spasial dan memiliki nilai intensitas antara 0 (hitam) hingga 255 (putih) dalam citra 8 bit (Putra, 2010). Khusus dalam library scikit-image direpresentasikan dengan bilangan float antara 0 sampai 1. Grayscale lebih mudah diproses karena mengandung nilai lebih sedikit, yaitu 8 bit warna dibandingkan RGB yang mengandung 24 bit warna. Untuk mengubah citra berwarna yang mempunyai nilai matrik masing- masing R, G, dan B menjadi citra grayscale dengan nilai s, maka konversi dapat dilakukan dengan mengambil rata-rata dari nilai r,g, dan b.

### Thresholding Background

Tahap ini dilakukan karena warna background tidak sepenuhnya putih. Terdapat variasi warna background yang sebagian besar intensitasnya mengarah ke warna putih sehingga agar seragam dan tidak mempengaruhi nilai komputasi maka dilakukan thresholding. Pada library scikit-image intensitas gambar grayscale direpresentasikan dengan rentang nilai antara 0 sampai 1. Pada studi kasus ini didapat nilai 0.75 sebagai threshold, sehingga piksel dengan intensitas lebih dari 0.75 akan dibulatkan menjadi satu.

## KLASIFIKASI K-NN

Pada tahap klasifikasi, data testing akan dibentuk matriks grayscale atau intensitas kemudian dibandingkan dengan seluruh data training untuk memeriksa data manakah yang lebih dekat dengan data yang akan di test tersebut. Prinsip dari K-NN adalah memeriksa apakah data testing tersebut dekat dengan sejumlah k tetangga terdekat. Jumlah kelas yang mempunyai lebih banyak kedekatan dengan data tersebut dianggap merupakan kelas yang tepat.

Penentuan kedekatan menggunakan L1 distance. Langkah awal dari tahap ini adalah dengan membentuk matriks Intensitas yang merupakan matriks 100x100. Langkah selanjutnya adalah menggunakan L1 distance untuk perbedaan piksel.

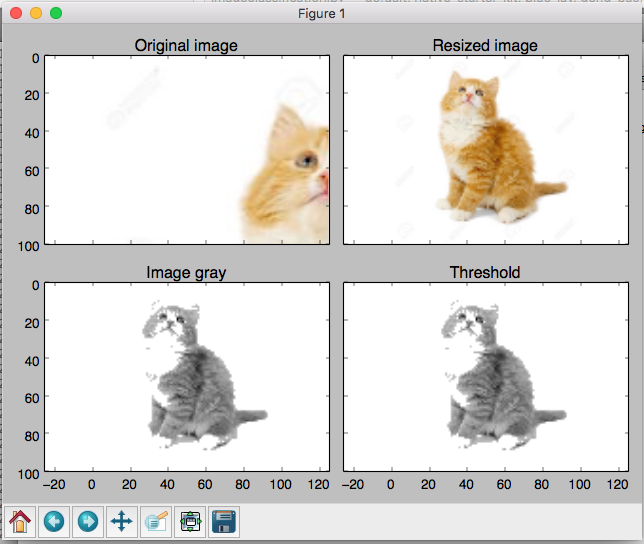
Tahap terakhir yaitu dengan melakukan sorting atau pengurutan terhadap data-data jarak yang telah dihitung berdasarkan jarak paling pendek. Berdasarkan data yang telah diurutkan tersebut dapat dilihat jumlah k tetangga paling dekat dari data tersebut untuk menentukan kelas.

# HASIL DAN PEMBAHASAN

## Contoh Pengambilan fitur data uji

Pengujian menggunakan nilai k= 5 pada metode pengujian dengan k-fold. Total data adalah 20 gambar dengan 10 gambar kucing dan 10 gambar anjing. Pengambilan fitur meliputi resizing image atau memperkecil gambar dengan ukuran 100x100. Kemudian diikuti oleh grayscaling dan thresholding nilai intensitas. Pada gambar dibawah ini akan disajikan contoh preprocessing sekaligus feature extraction.

Gambar awal merupakan gambar dengan ukuran 300x300 yang akan diperkecil menjadi 100x100 untuk mengurangi ruang pencarian. Kemudian dilakukan grayscaling dengan menghitung rata-rata rgb masing-masing piksel. Selanjutnya gambar terakhir merupakan hasil thresholding dimana nilai intensitas yang lebih dari 0.75 akan dibulatkan menjadi satu.



Gambar 1. Preprocessing dan Ekstraksi Fitur

## Metode Pengujian K-Fold dan Hasil Pengujian

Data uji yang digunakan berjumlah 4 buah dengan pembagian yang proporsional yaitu 2 gambar untuk kelas 1 dan 2 gambar untuk kelas 2. Sisanya yang berjumlah 16 digunakan sebagai data latih dengan pembagian yang proporsional juga yaitu 8 gambar untuk kelas 1 dan 8 gambar untuk kelas 2.

### Hasil pengujian K=1 dengan Ketetanggan Terdekat 5

DATA TESTING KE 0

label : cat

Daftar distance k terdekat :

dog=742.695274633

dog=766.654570506

cat=783.964364865

cat=784.913703704

cat=819.372092127

diklasifikasikan sebagai : cat

klasifikasi benar

DATA TESTING KE 1

label : dog

Daftar distance k terdekat :

dog=673.031870847

dog=673.149808845

dog=738.492113075

cat=769.666968494

dog=792.565214378

diklasifikasikan sebagai : dog

klasifikasi benar

DATA TESTING KE 2

label : cat

Daftar distance k terdekat :

cat=846.544341627

cat=849.238660412

cat=902.018546235

cat=985.342392235

cat=997.19559749

diklasifikasikan sebagai : cat

klasifikasi benar

DATA TESTING KE 3

label : dog

Daftar distance k terdekat :

dog=593.650586637

dog=683.010813733

dog=689.581662257

dog=706.816664727

dog=714.865161725

diklasifikasikan sebagai : dog

klasifikasi benar

Hasil pengujian dari k-fold untuk k bagian pertama sebagai data uji memperoleh akurasi 100%

### Hasil Pengujian K=2 dengan Ketetanggan Terdekat 5

DATA TESTING KE 0

label : cat

Daftar distance k terdekat :

cat=795.564435094

cat=849.238660412

cat=849.682993078

cat=857.946052953

cat=893.441050775

diklasifikasikan sebagai : cat

klasifikasi benar

DATA TESTING KE 1

label : dog

Daftar distance k terdekat :

dog=689.581662257

cat=715.515929635

dog=715.916314935

cat=742.695274633

dog=759.286672259

diklasifikasikan sebagai : dog

klasifikasi benar

DATA TESTING KE 2

label : cat

Daftar distance k terdekat :

cat=711.437324392

cat=889.875336863

cat=933.638126373

cat=947.007949022

cat=977.040968235

diklasifikasikan sebagai : cat

klasifikasi benar

DATA TESTING KE 3

label : dog

Daftar distance k terdekat :

dog=532.077123714

dog=532.66791972

dog=534.678259194

cat=602.17263001

dog=627.138969669

diklasifikasikan sebagai : dog

klasifikasi benar

Hasil pengujian dari k-fold untuk k bagian kedua sebagai data uji memperoleh akurasi 100%

### Hasil Pengujian K=3 dengan Ketetanggan Terdekat 5

DATA TESTING KE 0

label : cat

Daftar distance k terdekat :

cat=856.078258239

cat=889.875336863

cat=926.800211769

cat=993.904480575

cat=1051.04521529

diklasifikasikan sebagai : cat

klasifikasi benar

DATA TESTING KE 1

label : dog

Daftar distance k terdekat :

dog=534.678259194

dog=535.091384088

dog=549.802434933

cat=685.631501961

dog=722.3186095

diklasifikasikan sebagai : dog

klasifikasi benar

DATA TESTING KE 2

label : cat

Daftar distance k terdekat :

dog=715.515929635

cat=783.964364865

cat=798.755918227

dog=831.331622696

cat=833.246411275

diklasifikasikan sebagai : cat

klasifikasi benar

DATA TESTING KE 3

label : dog

Daftar distance k terdekat :

dog=552.311094508

dog=567.751035482

dog=593.650586637

dog=627.138969669

cat=732.467254463

diklasifikasikan sebagai : dog

klasifikasi benar

Hasil pengujian dari k-fold untuk k bagian ketiga sebagai data uji memperoleh akurasi 100%

### Hasil Pengujian K=4 dengan Ketetanggan Terdekat 5

DATA TESTING KE 0

label : cat

Daftar distance k terdekat :

cat=833.246411275

dog=882.230882255

dog=886.725090873

cat=902.018546235

cat=921.604187176

diklasifikasikan sebagai : cat

klasifikasi benar

DATA TESTING KE 1

label : dog

Daftar distance k terdekat :

dog=532.66791972

dog=549.802434933

dog=567.751035482

dog=673.031870847

dog=706.816664727

diklasifikasikan sebagai : dog

klasifikasi benar

DATA TESTING KE 2

label : cat

Daftar distance k terdekat :

dog=602.17263001

cat=668.022721745

dog=685.631501961

dog=732.467254463

dog=769.666968494

diklasifikasikan sebagai : dog

klasifikasi salah

DATA TESTING KE 3

label : dog

Daftar distance k terdekat :

dog=532.077123714

dog=535.091384088

dog=552.311094508

dog=673.149808845

cat=682.32380039

diklasifikasikan sebagai : dog

klasifikasi benar

Hasil pengujian dari k-fold untuk k bagian ketiga sebagai data uji memperoleh akurasi 75%

### Hasil Pengujian K=5 dengan Ketetanggan Terdekat 5

DATA TESTING KE 0

label : cat

Daftar distance k terdekat :

cat=668.022721745

dog=682.32380039

cat=711.437324392

dog=723.072553125

cat=784.913703704

diklasifikasikan sebagai : cat

klasifikasi benar

DATA TESTING KE 1

label : dog

Daftar distance k terdekat :

dog=823.705414902

dog=844.186554343

dog=898.542862494

dog=955.920697184

dog=963.216590308

diklasifikasikan sebagai : dog

klasifikasi benar

DATA TESTING KE 2

label : cat

Daftar distance k terdekat :

cat=1018.36451461

cat=1034.81185794

cat=1039.69809797

dog=1069.67577019

cat=1095.07826912

diklasifikasikan sebagai : cat

klasifikasi benar

DATA TESTING KE 3

label : dog

Daftar distance k terdekat :

dog=1525.18355327

dog=1543.07986412

dog=1570.80999547

cat=1575.86976711

cat=1579.39006828

diklasifikasikan sebagai : dog

klasifikasi benar

Hasil pengujian dari k-fold untuk k bagian ketiga sebagai data uji memperoleh akurasi 100%. Akurasi akhir dengan nilai ketetanggan 5 adalah 100+100+75+100/4 = 95%.

## EKSPERIMEN NILAI KETETANGGAN PADA KNN

|  |  |
| --- | --- |
| Nilai K | Akurasi (%) |
| 3 | 85 |
| 5 | 95 |
| 7 | 95 |
| 9 | 80 |
| 11 | 85 |
| 13 | 90 |
| 15 | 50 |

## PEMBAHASAN

Klasifikasi pada studi kasus ini menghasilkan akurasi terbaik yaitu 95% pada nilai k 5 dan 7 namun nilai k yang ideal merupakan sebuah knee atau lutut pada sebuah diagram perbandingan eksperimen nilai k. Jadi nilai k ideal dalam kasus ini adalah 9 dengan akurasi 80%.

**DAFTAR PUSTAKA**

Putra, Darma. 2010. Pengolahan Citra Digital. Yogyakarta : Penerbit Andi