Klasifikasi Neckband Sapi Menggunakan   
Metode Decision Tree, Random Forest,   
dan Logistic Regression

| Herdika Shidqi Wibowo1  G64190020  *Departemen Ilmu Komputer*  *Fakultas Matematika dan Ilmu Pengetahuan Alam*  IPB University, Bogor  Email: sw.herdika@apps.ipb.ac.id | Laudza Muhammad Afin Tachtiar2  G64190052  *Departemen Ilmu Komputer*  *Fakultas Matematika dan Ilmu Pengetahuan Alam*  IPB University, Bogor  Email: afintachtiar@apps.ipb.ac.id | Antonius Anre Sianturi3  G64190053  *Departemen Ilmu Komputer*  *Fakultas Matematika dan Ilmu Pengetahuan Alam*  IPB University, Bogor  Email: antoniusanre@apps.ipb.ac.id |
| --- | --- | --- |

| Abdul Hakim4  G64190078  *Departemen Ilmu Komputer*  *Fakultas Matematika dan Ilmu Pengetahuan Alam*  IPB University, Bogor  Email: hakimhakim@apps.ipb.ac.id | Farhan Fathurrahman5  G64190088  *Departemen Ilmu Komputer*  *Fakultas Matematika dan Ilmu Pengetahuan Alam*  IPB University, Bogor  Email: frhnfath\_farhan@apps.ipb.ac.id |
| --- | --- |

***Abstract*—Peternakan sudah dilakukan manusia sejak berabad-abad yang lalu untuk mendapatkan manfaat dari hewan yang diternakan. Memberikan kalung neckband pada sapi dapat dilakukan untuk mengidentifikasi jenis sapi yang diinginkan. Menggunakan algoritme feature extraction dan metode machine learning, dapat dilakukan identifikasi untuk membaca kalung neckband yang terdapat pada sapi. Metode feature extraction Histogram Oriented Gradient dan machine learning Decision Tree, Random Forest, dan Logistic Regression merupakan beberapa metode algoritma yang dapat digunakan untuk klasifikasi neckband tersebut. Dengan 1260 fitur yang dihasilkan metode HOG, hasil algoritme Decision Tree, Random Forest, dan Logistic Regression adalah model-model dengan akurasi klasifikasi 72.65%, 93.16%, dan 91.28%. Hasil akurasi menunjukkan Random Forest memiliki model terbaik untuk identifikasi neckband sapi.**

***Keywords*—*klasifikasi, decision tree, random forest, logistic regression, neckband sapi, pengolahan citra***

# I. PENDAHULUAN

Pada zaman modern yang semakin canggih ini, manusia tentu akan seberusaha mungkin untuk menciptakan suatu penemuan yang bermanfaat dan dapat memudahkan pekerjaan dunia sehari-hari. Di segala aspek kehidupan, peran ilmu pengetahuan dan teknologi telah terasa sangat nyata adanya. Teknologi tidak hanya berguna untuk diterapkan pada bidang telekomunikasi, tetapi banyak juga teknologi yang bisa diterapkan pada peternakan, pertanian, kelautan, dan sebagainya.

Seiring berkembangnya waktu, kebutuhan manusia akan pangan semakin meningkat. Sektor pertanian dan peternakan harus dapat menyesuaikan perkembangan zaman dengan lebih mengedepankan perkembangan ilmu pengetahuan dan teknologi apabila ingin bersaing di pasar global. Bahkan menurut laporan Dirjen Peternakan di tahun 2012, disebutkan bahwa nilai PDB subsektor peternakan pada tahun 2011 memiliki dasar harga konstan sebesar Rp 313,7 triliun, menaik sebesar 2,95% dibanding dengan tahun 2010. Nilai ini akan diprediksi terus naik di setiap tahunnya [1]. Angka tersebut merupakan indikasi bahwa sektor peternakan di Indonesia tidak dapat dipandang sebelah mata, apalagi jika mengingat bahwa negara ini merupakan negara agraris.

Peternakan sendiri merupakan salah satu bidang penting yang menyokong ketahanan pangan Indonesia. Industri peternakan yang besar memiliki hewan yang sangat banyak. Tidak semua hewan ternak memiliki karakteristik yang sama. Sangat sulit untuk mengidentifikasi kelompok hewan ternak apabila seandainya hewan-hewan tersebut dilepaskandangkan. Pada kasus tersebut, teknologi dapat dimanfaatkan dan diimplementasikan untuk mengidentifikasi hewan ternak dengan tujuan memudahkan dalam mengelompokkan kategori ataupun jenis hewan-hewan ternak tersebut.

Salah satu pemanfaatan teknologi pada peternakan dapat diimplementasikan pada peternakan sapi. Penerapannya adalah dengan memberikan sapi sejenis *neckband* (gelang yang dipasangkan di leher) yang sesuai dengan kategori dan jenis sapi untuk memudahkan klasifikasi sapi-sapi tersebut. Dengan adanya *neckband* ini, sapi-sapi akan dapat lebih mudah dikenali kelompok dan jenisnya. Pada penerapan lebih lanjut, identifikasi sapi dengan aksesoris tambahan ini dapat membantu pengenalan jenis sapi melalui media identifikasi objek dengan menggunakan sensor seperti kamera.

Tujuan dilakukannya penelitian ini adalah untuk membuat model *machine learning* yang dapat mengidentifikasi objek citra *neckband* pada sapi dengan tujuan untuk mengklasifikasikan jenis dan kategori sapi berdasarkan bentuk simbol pada penampang *neckband* yang dikenakan oleh sapi. Teknik yang akan digunakan dalam proses pendeteksian objek *neckband* sapi adalah HOG (Histogram of Oriented Gradient) Edge Detection. Teknik identifikasi objek tersebut akan dipadukan dengan algoritme *machine learning* yang digunakan pada proses identifikasi citra hasil HOG. Algoritme yang digunakan pada penelitian ini adalah Decision tree, Random forest, dan Logistic regression. Penerapan ketiga model yang berbeda ini diharapkan dapat dibuat perbandingan hasil dan juga nilai kebaikan modelnya.

# II. TINJAUAN PUSTAKA

## A. Supervised Learning

Supervised learning atau dalam bahasa Indonesia berarti pembelajaran yang disupervisi merupakan bagian dari teknik machine learning yang mempelajari fungsi yang akan memetakan satu atau lebih variabel prediktor ke satu variabel target berdasarkan contoh pasangan input-output yang telah disediakan sebelumnya [2]. Hal tersebut berimplikasi bahwa algoritme supervised learning ini terdiri dari dataset yang memiliki label yang nantinya akan dijadikan sebuah set contoh data latih. Dalam praktiknya, pengembangan model supervised learning akan membagi data menjadi data latih dan data tes untuk menguji seberapa baik model tersebut dalam memprediksi suatu data baru di luar data yang digunakan dalam pembuatan model [3].

## B. Klasifikasi Citra

Proses klasifikasi dengan media input-output yang direpresentasikan oleh gambar atau citra memiliki perlakuan dan karakteristik khusus. Hal yang membedakan antara proses pembuatan model supervised learning antara data yang berbentuk tabular dan citra adalah banyaknya jumlah fitur yang terdapat dalam citra. Banyaknya fitur atau atribut yang mengandung informasi dalam citra berupa dimensi piksel dari citra itu sendiri. Tentu hal ini menjadi masalah apabila data yang digunakan tidak melalui proses penyederhanaan dan pemilihan model yang cocok. Dalam prosesnya, algoritme deep learning biasanya dipilih karena memiliki aspek ekstraksi fitur yang secara langsung terkandung di dalam modelnya [4]. Algoritme machine learning lainnya juga dapat diterapkan untuk klasifikasi citra, akan tetapi proses ekstraksi fitur yang diterapkan biasanya dilakukan di luar model algoritme yang dibangun [5].

## C. Histogram of Oriented Gradients

Pada beberapa tahun terakhir, algoritme HOG (Histogram of Oriented Gradients) semakin populer. Banyak peneliti yang menggunakan algoritme ini untuk mendeteksi suatu objek pada citra dan memperoleh akurasi yang baik. Metode ini memiliki tujuan untuk mengenali suatu objek dengan histogram yang merepresentasikan frekuensi nilai piksel pada blok tertentu. Multi histogram ini melibatkan nilai-nilai gradien lokal pada area blok gambar yang mengarah ke luar. Realisasi dari proses HOG ini adalah sebagai berikut: 1). Penerimaan nilai gradien dari suatu citra; 2). Penetapan arah histogram pada lokasi atau area tertentu; dan 3). Normalisasi histogram pada lokasi blok yang spesifik [6].

Algoritme HOG adalah jenis dari “quality descriptor” yang akan menggeneralisasi karakteristik objek terdeteksi sedemikian rupa sehingga objek yang sama akan teridentifikasi sebagai objek yang serupa walau dalam keadaan blok sekitar yang berbeda-beda.

## D. Logistic Regression

Regresi logistik atau bisa disebut juga model logit adalah suatu model regresi yang digunakan untuk memprediksi probabilitas kejadian suatu peristiwa dengan mencocokkan data pada fungsi logit kurva logistik. Sebagaimana regresi linear, metode ini menggunakan beberapa variabel prediktor, baik numerik maupun kategorik [7].

## E. Decision Tree

Secara sederhana, algoritme Decision Tree dapat disebut sebagai analisis klasifikasi yang dilakukan dengan pendekatan prinsip divide-and-conquer. Algoritme ini dapat digunakan untuk mengetahui tingkat kepentingan suatu fitur dan model pola diskriminasi kelas sekumpulan objek berdasarkan parameter tertentu. Parameter umum yang digunakan pada penentuan importance feature tersebut biasanya adalah entropi, gini, dan information gain [8].

## F. Random Forest

Random Forest merupakan algoritma yang memiliki ide dasar untuk membuat banyak Decision Tree pada saat proses pelatihan. Untuk tujuan klasifikasi, hasil output dari pemodelan ini adalah kelas yang terpilih oleh tree terbanyak. Di sisi lain, untuk kepentingan regresi, output yang diperoleh adalah nilai rerata dari setiap tree yang dihasilkan. Secara umum, akurasi kebaikan dari model Random Forest akan lebih baik daripada Decision Tree [9].

# III. METODE

## A. Augmentasi

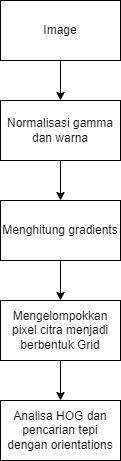
Augmentasi merupakan proses mengolah data citra dengan cara memodifikasi gambar menjadi sedemikian rupa sehingga akan menghasilkan citra yang dapat dikenali oleh komputer sebagai gambar yang berbeda. Tahapan augmentasi yang dilakukan adalah rotasi, scaling, mengatur kecerahan, horizontal dan vertical flip, translasi dan shearing.

Setelah dilakukan pembagian dataset menjadi data latih dan data uji, dilakukan pemanggilan fungsi augmentasi ke dalam data latih sehingga dari data 1 citra tersebut akan diolah dan didapatkan 6 data citra lainnya melalui proses augmentasi yang sudah dilakukan. Data baru yang didapatkan adalah data citra yang dilakukan rotasi, scaling, diatur kecerahan, horizontal dan vertical flip, translasi dan shearing. Sehingga total didapatkan data latih yang 7 kali lebih besar dari data latih awal.

Augmentasi dilakukan untuk mensimulasikan pengambilan data yang tidak sempurna, seperti kecerahan yang tidak baik, jarak pengambilan citra yang kurang tepat, dan kondisi-kondisi tidak ideal lainnya. Diharapkan dengan adanya penambahan beberapa citra hasil augmentasi akan meningkatkan akurasi dan kemampuan model dalam mengidentifikasi pola neckband pada sapi.

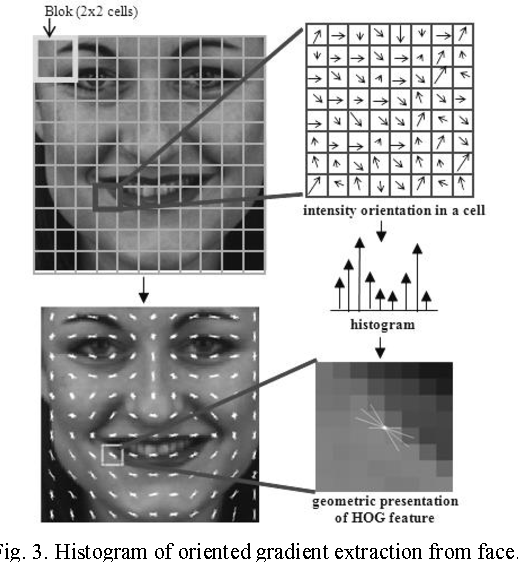
## B. Histogram of Oriented Gradient (HOG)

Ekstraksi fitur menggunakan HOG diterapkan kepada data latih dan data uji, terdapat beberapa parameter yang terdapat pada fungsi hog yaitu orientations berguna untuk seberapa banyak arah yang digunakan untuk mendeteksi tepian yang terdapat pada citra. Pixel\_per\_cell membagi pixel citra ke dalam jumlah cell yang diinginkan. Cell\_per\_block mengelompokkan cell ke dalam 1 *block* berdasarkan besaran yang diinginkan.



Gambar 1: Flowchart proses HOG

Proses ekstraksi fitur pada HOG akan mengelompokkan pixel ke dalam cell dan kemudian mengelompokkan cell ke dalam suatu block menjadi berbentuk Grid. Analisa histogram akan dilakukan untuk mencari tepian (edge) yang terdapat pada citra, ditambah dengan parameter *orientations* sebagai jumlah arah untuk menentukan *edge detection* yang ada.



Gambar 2: Contoh proses HOG [10]

## C. Logistic Regression

Pemodelan menggunakan *Logistic Regression* dilakukan dengan melakukan scaling terhadap data latih, kemudian melakukan *fitting model* untuk mengimplementasikan algoritma logistic regression ke dalam data latih yang sudah ada.

Setelah selesai melakukan *fitting model*, maka melakukan prediksi dengan data uji untuk mengetahui performa model dan dianalisis mengenai akurasi model yang sudah dibuat sebelumnya.

Pada fungsi yang digunakan hanya memodifikasi parameter random\_state = 1, parameter lainnya menggunakan parameter default dari fungsi scikit-learn. Terdapat beberapa parameter yang dapat dimodifikasi untuk mendapatkan model *logistic regression* yang optimal, seperti :

**Class\_weight**, untuk inisiasi bobot

**Random\_state**, konstanta untuk standar random

**Max\_iter**, maksimum iterasi yang dilakukan

## D. Decision Tree

Pada pemodelan menggunakan Decision Tree yang diterapkan pada upaya klasifikasi objek *neckband* sapi yang sebelumnya telah diekstraksi menggunakan teknik HOG, *library* dan fungsi yang digunakan untuk memudahkan dalam komputasi adalah sklearn.DecisionTreeClassifier.

Pada fungsi yang telah disebutkan di atas, terdapat berbagai macam parameter yang digunakan untuk kustomisasi bentuk dan perhitungan dari model Decision Tree yang akan dibuat. Parameter-parameter tersebut antara lain adalah:

**criterion**, fungsi yang menentukan pembagian cabang.

**splitter**, strategi yang digunakan untuk membagi cabang setiap *node*.

**max\_depth**, maksimum kedalaman Decision Tree. Semakin tinggi nilai ini maka semakin kompleks dan akurat model tersebut.

**min\_samples\_split**, minimum banyaknya sampel yang dibutuhkan untuk memecah suatu *node*.

**min\_samples\_split**, minimum banyaknya sampel yang dibutuhkan untuk menjadi suatu *leaf node*.

**min\_weight\_fraction\_leaf**, minimum bobot dari fraksi node untuk menjadi suatu *leaf node*.

**max\_features**, banyaknya fitur yang akan dipertimbangkan dalam proses pemecahan cabang.

**random\_state**, konstanta yang menentukan standar pengacakan.

**min\_impurity\_decrease**, nilai yang ditetapkan sebagai *impurity threshold* pada setiap percabangan node.

**ccp\_alpha**, nilai kompleksitas yang digunakan dalam upaya meminimumkan Cost-Complexity Pruning.

Pada pemodelan menggunakan Decision Tree nanti, parameter-parameter di atas akan diatur sedemikian rupa agar nilai kebaikan model berdasar data latih yang digunakan menjadi lebih optimal.

## E. Random Forest

Pada pemodelan menggunakan Random Forest yang diterapkan pada upaya klasifikasi objek *neckband* sapi yang sebelumnya telah diekstraksi menggunakan teknik HOG, *library* dan fungsi yang digunakan untuk memudahkan komputasi ialah sklearn.ensemble.RandomForestClassifier.

Seperti yang telah diterangkan pada bagian tinjauan pustaka, karena Random Forest merupakan kumpulan dari banyak Decision Tree dengan struktur yang berbeda-beda, maka parameter-parameter yang digunakan pada pemodelan menggunakan algoritme ini tidaklah jauh berbeda. Terdapat beberapa parameter tambahan yang dikhususkan untuk Random Forest, di antaranya adalah:

**n\_estimators**, banyaknya *tree* dari algoritme Decision Tree yang akan di-*generate*.

**bootstrap**, mengaktifkan ansambel dengan membagi data latih sebelum membangun *tree*. Hal ini dilakukan dengan tujuan membuat model semakin *robust*.

**oob\_score**, menentukan penggunaan sampel *out-of-bag* untuk mengestimasi skor ketika *bootstrap* bernilai True.

**n\_jobs**, banyaknya proses pembuatan *tree* yang bekerja secara paralel.

**verbose**, nilai kontrol yang digunakan untuk mengatur verbositas pada saat proses pengepasan dan peramalan.

**warm\_start**, penggunaan kembali solusi atau *tree* yang telah dibuat sebelumnya ketika membangun model baru dengan parameter yang sama.

Pada pemodelan menggunakan Random Forest nanti, parameter-parameter di atas akan diatur sedemikian rupa agar nilai kebaikan model berdasar data latih yang digunakan menjadi lebih optimal.

# IV. HASIL DAN PEMBAHASAN

## A. Data Neckband Sapi

Data neckband sapi yang diperoleh terbagi menjadi delapan kelas yang dibedakan berdasarkan gambar simbol yang terdapat pada neckband sapi. Simbol berbentuk bulatan dengan pola berwarna hitam dan putih yang terletak beberapa di keliling neckband. Neckband diposisikan pada leher sapi dan gambar diambil saat sapi menghadap ke bawah. Berikut adalah contoh gambar yang diambil untuk setiap kelas;

TABEL I: Data citra neckband sapi

| (a) Class 1 | (b) Class 2 |
| --- | --- |
| (c) Class 3 | (d) Class 4 |
| (e) Class 5 | (f) Class 6 |
| (g) Class 7 | (h) Class 8 |

Terlihat pada gambar di atas bahwa simbol-simbol yang digunakan untuk membedakan kelas neckband sapi. Neckband pada sapi digunakan sebagai untuk membedakan 8 ekor sapi yang diletakkan pada kandang yang sama. Simbol kelas semua berbentuk lingkaran dengan garis luar berwarna hitam yang tebal dengan pola yang berbeda untuk masing-masing kelas. Pola simbol dapat dilihat lebih jelas pada tabel di bawah.

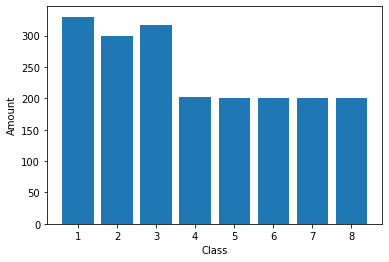
TABEL II: Simbol neckband sapi

| Class 1 | Class 2 | Class 3 | Class 4 |
| --- | --- | --- | --- |
| Class 5 | Class 6 | Class 7 | Class 8 |

Data citra neckband untuk semua class akan digunakan untuk penelitian ini. Jumlah citra pada setiap class adalah seperti berikut.

TABEL III: Jumlah data per kelas

| **Class** | **Jumlah Data** |
| --- | --- |
| 1 | 330 |
| 2 | 299 |
| 3 | 316 |
| 4 | 202 |
| 5 | 200 |
| 6 | 200 |
| 7 | 200 |
| 8 | 201 |
| **Total** | **1948** |



Gambar 3: Distribusi jumlah data per kelas

Gambar di atas merupakan sebaran data untuk setiap class. Class 1, class 2 dan class 3 memiliki jumlah kelas yang lebih banyak dibanding lainnya yang rata-rata mendekati total 200 citra.

## B. Identifikasi Neckband Tak Terlihat

Klasifikasi neckband sapi memerlukan citra yang memperlihatkan neckband dengan jelas. Dari data yang kami dapatkan, terdapat beberapa citra dari setiap kelas yang tidak terlihat neckbandnya sehingga harus dipisah dari class-nya. Alasan tidak terlihat neckband karena tertutup dengan pagar yang memisah badan sapi dengan tempat makan sapi ataupun neckbandnya yang hanya terihat sebagian. Berikut adalah beberapa contoh citra tersebut.

TABEL IV: Contoh neckband sapi tak terlihat

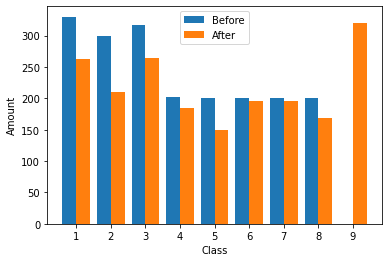
|  |  |
| --- | --- |
|  |  |

Penanganan yang dilakukan adalah dengan memisah citra yang tak terlihat menjadi class baru. Pendekatan ini dilakukan, dan bukan penghapusan, karena ingin dapat mengidentifikasi jenis citra seperti itu agar dapat mengurangi kesalahan saat klasifikasi dan lebih ramah terhadap data aktual.

Citra dengan neckband pada setiap kelas dimasukkan dalam kelas baru yang akan diberi nama “Class 9”. Hal tersebut akan mengurangi jumlah citra pada setiap class dengan masing-masing jumlah seperti berikut.

TABEL V: Jumlah data setelah citra tak terlihat dipisahkan

| **Class** | **Jumlah Data** | **Tak Terlihat** |
| --- | --- | --- |
| 1 | 262 | 68 |
| 2 | 210 | 89 |
| 3 | 264 | 52 |
| 4 | 184 | 18 |
| 5 | 149 | 51 |
| 6 | 195 | 5 |
| 7 | 195 | 5 |
| 8 | 169 | 32 |
| **Total** | **1628** | **320** |

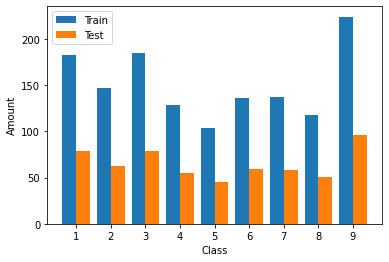


Gambar 4: Distribusi jumlah data per kelas sebelum dan sesudah pemisahan Kelas 9

Class 2 memiliki citra dengan neckband tak terlihat yang terbanyak sebesar 89 citra dan Class 6 serta Class 7 yang paling sedikit dengan hanya 5 citra. Setelah penelusuran dan pencarian citra tak terlihat pada setiap kelas, terakumulasi sebesar 320 citra, atau 16.4% dari total citra, yang akan dimasukkan pada Class 9.

## C. Data Latih dan Data Uji

Citra yang diperoleh dipartisi menjadi data latih dan data uji masing-masing sebesar 70% dan 30%. Jumlah masing-masing data latih dan data uji adalah 1363 citra dan 585 citra.



Gambar 5: Distribusi data latih dan data uji per kelas

Data latih dan data uji yang ditarik memiliki distribusi yang serupa dengan jumlah data citra awal agar proses klasifikasi lebih baik. Kelas 9 memiliki nilai tertinggi sedangkan Kelas 1 dan 3 masih cukup tinggi dan kelas lainnya sedang.

## D. Augmentasi Citra

Augmentasi citra dilakukan pada setiap citra dalam setiap kelas jumlah citra pada data latih agar citra yang digunakan untuk klasifikasi akan menambah. Proses augmentasi yang digunakan adalah;

1. Rotation : Antara -180° dan 180°
2. Scaling : Antara x1 dan x3
3. Brightness : Kecerahan antara 0 hingga 1
4. Horizontal Flip : Balik horizontal
5. Vertical Flip : Balik vertikal
6. Translation : Antara 0 hingga ½ tinggi dan antara 0 hingga ½ lebar.
7. Shearing : Antara -45 dan 45

Nilai yang diberikan pada setiap augmentasi diberi secara acak agar memberi citra input yang lebih bervariasi. Namun parameter yang diberikan tidak terlalu ekstrem agar gambar tetap bisa dilihat dan diproses. Berikut adalah contoh output dari augmentasi yang dilakukan;

TABEL VI: Contoh hasil augmentasi citra

| **Augmentasi** | **Hasil** |
| --- | --- |
| Rotation |  |
| Scaling |  |
| Kecerahan |  |
| Horizontal Flip |  |
| Vertical Flip |  |
| Translation |  |
| Shearing |  |

Setelah pengimplementasian augmentasi pada citra, didapatkan jumlah citra masing-masing kelas seperti pada tabel di bawah.

TABEL VII: Jumlah data latih setelah augmentasi

| **Class** | **Jumlah Data Latih** |
| --- | --- |
| 1 | 1464 |
| 2 | 1176 |
| 3 | 1480 |
| 4 | 1032 |
| 5 | 832 |
| 6 | 1088 |
| 7 | 1096 |
| 8 | 944 |
| 9 | 1792 |
| **Total** | **10904** |

Jumlah data latih menaik dari 1363 menjadi 10904 yaitu tepat 8 kali dari jumlah data latih awal. Augmentasi data latih akan membuat pembelajaran pengklasifikasian lebih baik, bukan hanya dari jumlah data, namun juga dari tujuan augmentasi yaitu untuk memberi proses klasifikasi lebih banyak variasi citra.

## E. Histogram of Oriented Gradient (HOG)

Pada proses implementasi Histogram of Oriented Gradient (HOG), *,* menggunakan fungsi hog dari skimage.feature dan didapatkan hasil yang optimal dengan parameter sebagai berikut :

TABEL VIII: Parameter HOG

| **Parameter** | **Nilai** |
| --- | --- |
| Orientations | 9 |
| pixel\_per\_cell | (8,8) |
| cells\_per\_block | (2,2) |

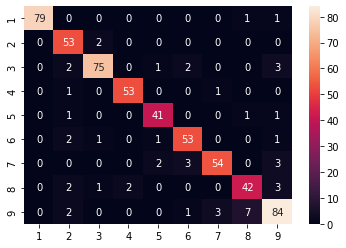
Setelah menggunakan parameter berikut, didapatkan ekstraksi fitur untuk mencari *edge* yang terdapat pada citra, dengan contoh salah satu hasil citra sebagai berikut

  
Gambar 6: Hasil HOG

Terlihat pada citra diatas pada vector orientations mampu mendeteksi tepian yang kemudian fiturnya akan digunakan untuk attributes model*.* Hasil pengolahan ekstraksi fitur tersebut kemudian disimpan ke dalam variable. Diperoleh hog features sebanyak **1260** yang dapat digunakan sebagai *features* untuk pengolahan model *machine learning*.

## F. Logistic Regression

Menggunakan fungsi LogisticRegression dari sklearn, dilakukan modifikasi parameter random\_state = 1 untuk menentukan konstanta standar random. Parameter lainnya menggunakan nilai default yang telah disediakan oleh fungsi tersebut. Berikut visualisasi *confusion matrix* yang dihasilkan oleh model Logistic Regression :



Gambar 7: Confusion matrix Logistic Regression

Terlihat bahwa model yang dihasilkan sudah sangat baik dalam klasifikasi citra. Melihat kelas 1, terlihat bahwa seluruh citra dapat diidentifikasi dengan benar. Pada class 9, terdapat 84 citra dapat diidentifikasi dengan benar, dan hanya 12 citra yang salah diidentifikasi. Terlihat juga pada kelas lainnya. Pada kelas 2 dan 8 memiliki persentase akurasi yang cukup rendah dibandingkan kelas lainnya dengan akurasi 84.13% dan 87.5%. Nilai akurasi keseluruhan yang dihasilkan model *Logistic Regression* adalah sebesar 91.28%. Melihat dari *confusion matrix* diatas dan hasil akurasi yang didapatkan, maka dapat disimpulkan bahwa hasil model *Logistic Regression* sangat baik dalam mengidentifikasi citra *neckband* pada sapi.

## G. Decision Tree

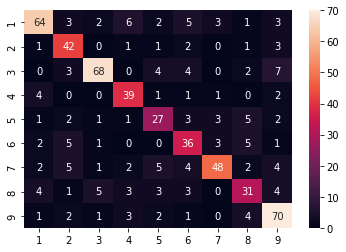
Pada proses klasifikasi menggunakan algoritme Decision Tree dengan fungsi DecisionTreeClassifier(), parameter-parameter yang diterapkan untuk menghasilkan hasil yang mendekati optimal tersaji pada tabel berikut.

TABEL IX: Parameter Decision Tree

| **Parameter** | **Nilai** |
| --- | --- |
| criterion | gini |
| splitter | best |
| max\_depth | *None* |
| min\_samples\_split | 2 |
| min\_samples\_leaf | 1 |
| min\_weight\_fraction\_leaf | 0.0 |
| max\_features | *None* |
| random\_state | 1 |
| min\_impurity\_decrease | 0.0 |
| ccp\_alpha | 0.0 |

Setelah fungsi DecisionTreeClassifier dijalankan dengan kombinasi parameter seperti yang tertera pada tabel di atas, model yang dihasilkan langsung diujicobakan dengan data tes yang sebelumnya juga telah melalui fitur ekstraksi HOG. Setelah dilakukan perbandingan label pada tiap data tes, didapatkan nilai dari akurasi model tersebut adalah sebesar **72.65%**.

Berikut ditampilkan visualisasi *confusion matrix* dari pemodelan menggunakan Decision Tree.



Gambar 8: Confusion matrix Decision Tree

Pada confusion matrix di atas masih terlihat bahwa terdapat cukup banyak data atau citra yang salah diidentifikasi. Sebagai contoh, pada Class 9, dari jumlah sampel data tes sebanyak 96 citra, hanya 70 di antaranya yang berhasil diidentifikasi sebagai Class 9. 26 cita lainnya masih salah diidentifikasi. Bahkan jika dilihat secara keseluruhan, tidak ada satu pun kelas yang memiliki keakuratan identifikasi sebesar 95% ke atas.

## Berdasarkan dari informasi yang tersaji, dapat dikatakan bahwa pemodelan menggunakan Decision Tree dalam mengidentifikasi kelas berdasarkan fitur-fitur HOG dari citra neckband sapi masih menunjukkan performa yang kurang baik. Hal ini dapat dimaklumi karena pemodelan menggunakan Decision Tree masih dapat digolongkan sebagai pemodelan machine learning yang relatif sederhana dan hanya dapat memiliki akurasi yang baik untuk kasus yang lebih sederhana.

## H. Random Forest

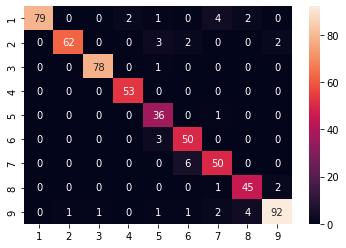
Pada proses klasifikasi menggunakan algoritme Decision Tree dengan fungsi RandomForestClassifier(), parameter-parameter yang diterapkan untuk menghasilkan hasil yang mendekati optimal tersaji pada tabel berikut.

TABEL XI: Parameter Random Forest

| **Parameter** | **Nilai** |
| --- | --- |
| n\_estimators | 100 |
| criterion | gini |
| max\_depth | *None* |
| min\_samples\_split | 2 |
| min\_samples\_leaf | 1 |
| min\_weight\_fraction\_leaf | 0.0 |
| max\_features | *sqrt* |
| random\_state | 1 |
| min\_impurity\_decrease | 0.0 |
| ccp\_alpha | 0.0 |
| bootstrap | True |
| oob\_score | False |
| n\_jobs | *None* |
| verbose | 0 |

Setelah fungsi RandomForestClassifier dijalankan dengan kombinasi parameter seperti yang tertera pada tabel di atas, model yang dihasilkan langsung diujicobakan dengan data tes yang sebelumnya juga telah melalui fitur ekstraksi HOG. Setelah dilakukan perbandingan label pada tiap data tes, didapatkan nilai dari akurasi model tersebut adalah sebesar **93.16%**.

Berikut ditampilkan visualisasi *confusion matrix* dari pemodelan menggunakan Random Forest.



Gambar 9: Confusion matrix Random Forest

## Pada confusion matrix di atas terlihat bahwa model yang diterapkan pada proses identifikasi citra telah menunjukkan performa yang jauh lebih baik daripada model Decision Tree. Sebagai contoh perbandingan, pada Class 9, dari 96 sampel yang ada, 92 di antaranya berhasil diidentifikasi sebagai objek citra Class 9. Pada model Decision Tree sebelumnya, hanya 70 data saja yang dapat diidentifikasi dengan benar. Hal ini menunjukkan peningkatan akurasi yang cukup signifikan antara pemodelan dengan menggunakan Decision Tree dan juga Random Forest.

Sejalan dengan kelas lainnya, terlihat bahwa pada *Class 1*, seluruh data citra dapat diidentifikasi dengan akurat. Begitupun dengan kelas-kelas lainnya yang dimana nilai akurasinya juga relatif tinggi. Terlihat hanya *Class 5*, *Class 6*, dan *Class 7* sajalah yang masih memiliki akurasi di bawah angka 90%.

Peningkatan akurasi ini begitu menggambarkan bagaimana *powerful*-nya algortime Random Forest jika dibandingkan Decision Tree, karena memang pada dasarnya Random Forest merupakan kumpulan dari banyak *tree* dari Decison Tree yang memiliki struktur yang berbeda-beda. Sehingga model yang dihasilkan akan cenderung lebih *robust* serta memiliki akurasi yang lebih tinggi..

## I. Perbandingan Algoritme Klasifikasi

Ketiga algoritme yang sudah dijalankan dibandingkan hasilnya. Akurasi dari ketiga algoritme tersebut dapat menunjukkan performa dari masing-masing algoritme terhadap kasus identifikasi neckband sapi ini. Jadi, ketiga akurasi dapat dibandingkan untuk menentukan algoritme mana yang paling baik untuk identifikasi *neckband* sapi.

Perbandingan akurasi ketiga algoritme adalah dalam tabel seperti berikut.

TABEL XII: Perbandingan akurasi hasil klasifikasi

| **Algoritme** | **Akurasi** |
| --- | --- |
| Logistic Regression | 91.28% |
| Decision Tree | 72.65% |
| Random Forest | 93.16% |

Besarnya akurasi untuk setiap algoritme tidak terlepas dari karakteristik dari algoritme itu sendiri. Algoritme decision tree yang menggunakan pohon keputusan dalam bentuk rule menghasilkan model dengan akurasi terendah bila dibandingkan dengan dua algoritme lainnya. Model decision tree yang kurang baik ini dikarenakan data yang dihasilkan dari ekstraksi fitur pada citra tentunya memiliki pola yang sangat bervariatif. Variasi ini membuat tidak semua citra memiliki fitur utama yang sama, sedangkan pada algoritme decision tree menggunakan fitur utama yang sama untuk klasifikasi setiap citranya.

Model klasifikasi yang menggunakan algoritme logistic regression merupakan model terbaik kedua berdasarkan besar akurasinya. Logistic regression mampu memberikan hasil yang lebih baik dibandingkan decision tree karena logistic regression menggunakan fungsi untuk menghitung peluang setiap citra yang parameternya terdiri dari beberapa fitur penting dengan bobot berbeda-beda. Bobot dan beberapa parameter yang digunakan membuat algoritme ini mampu menangani lebih banyak variasi data.

Model terbaik berdasarkan akurasinya adalah model yang menggunakan algoritme random forest. Algoritme random forest sendiri merupakan peningkatan dari algoritme decision tree. Pada model ini, rule ataupun aturan yang dibentuk tidak harus memiliki satu fitur utama yang sama, melainkan dapat terklasifikasi menggunakan beberapa kombinasi fitur yang berbeda dan nantinya akan diambil kombinasi fitur ataupun rule yang terbaik. Fleksibilitas dari algoritme random forest membuat model mampu menangani berbagai variasi data dan menghasilkan akurasi terbaik.

# V. SIMPUL DAN SARAN

Sapi yang berada di kandang berbeda diberikan *neckband* atau kalung pada leher yang memiliki pola beragam. Pola- pola tersebut akan digunakan sebagai pembeda kelompok atau koloni dan jenis sapi sehingga akan lebih mudah dalam mengidentifikasi seekor sapi. Melalui pengolahan citra digital, komputer ataupun sistem bisa membedakan *neckband* yang beragam tersebut. Pada penelitian kali ini, diterapkan 3 algoritme klasifikasi, Decision Tree, Logistic Regression, dan Random Forest, untuk membuat model pembeda antar-neckband. Secara berurutan, ketiga algoritme tersebut menghasilkan nilai akurasi sebesar 72.65%, 91.28%, dan 93.16%. Berdasarkan nilai akurasi dari ketiga algoritme tersebut, dapat dinyatakan bahwa algoritme Random Forest memiliki performa paling baik untuk klasifikasi citra sapi yang menggunakan *neckband*. Nilai akurasi yang diperoleh Terdapat beberapa saran yang bisa digunakan untuk penelitian ke depannya, antara lain :

* Parameter yang digunakan untuk membangun model ditentukan melalui proses hyperparameter tuning dengan tujuan meningkatkan kebaikan model.
* Pada tahapan pra-proses data dapat dilakukan segmentasi terlebih dahulu pada bagian neckband guna meningkatkan kebaikan model.

# REFERENSI

1. S. Nurtini, M. Anggriani. *Profil Peternakan Sapi Perah Rakyat di Indonesia*. Yogyakarta: UGM PRESS, 2018.
2. S. J. Russell, P. Norvig. *Artificial Intelligence: A Modern Approach, Third Edition*. New Jersey: Prentice Hall, 2010.
3. M. Mohri, A. Rostamizadeh, A. Talwalkar. *Foundations of Machine Learning*. Boston: The MIT Press, 2012.
4. D. Lu, Q. Weng. “A survey of image classification methods and techniques for improving classification performance,” *International Journal of Remote Testing*, vol. 28, pp. 823-870, March 2007.
5. B. Kumar, O. Dikshit, A. Gupta, M. K. Singh. “Feature extraction for hyperspectral image classification: a review,” *International Journal of Remote Testing*, vol. 41, pp. 6248-6287, June 2020.
6. P. Wang E. Fan P. Wang. “Comparative analysis of image classification algorithms based on traditional machine learning and deep learning,” *Pattern Recognition Letters,* vol. 141, pp. 61-67, January 2021.
7. D. G. Kleinbaum, M. Klein. *Logistic Regression: A Self-Learning Text*. New York: Springer, 2010.
8. A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, S. D. Brown. “An introduction to decision tree modeling,” *Journal of Chemometrics*, vol. 18, pp. 275-285, June 2004.
9. Tin Kam Ho, "Random decision forests," *Proceedings of 3rd International Conference on Document Analysis and Recognition*, pp. 278-282 vol.1, August 1995.
10. L. Greche, N. Es-Sbai, "Automatic system for facial expression recognition based histogram of oriented gradient and normalized cross correlation," *2016* *International Conference on Information Technology for Organizations Development*, pp. 1-5, March 2016.