

PERBANDINGAN METODE NAÏVE BAYES DAN C4.5 UNTUK MEMPREDIKSI MORTALITAS PADA PETERNAKAN AYAM BROILER

Dimas Imam Baihaqi

Fakultas Teknik, Program Studi Teknik Informatika
Universitas Negeri Malang
Email: dimas.bai9@gmail.com

Anik Nur Handayani

Jurusan Elektro
Universitas Negeri Malang
Email: anik.nur.ft@um.ac.id

Utomo Pujiyanto

Jurusan Elektro
Universitas Negeri Malang
Email: utomo.pujiyanto.ft@um.ac.id

ABSTRAK

Ayam broiler adalah jenis ternak yang paling cepat untuk dipanen. Namun dalam berternak ayam broiler pasti banyak masalah yang dihadapi contohnya adalah tingkat kematian. Untuk menekan kerugian, para peternak sebaiknya memperhatikan faktor-faktor apa saja yang menyebabkan kematian ayam tersebut. Beberapa penelitian yang meneliti tentang ayam broiler menggunakan metode percobaan dan RAL. Namun masih belum ada yang meneliti mortalitas ayam broiler menggunakan komputasi. Untuk mengetahui metode mana yang lebih baik untuk memprediksi mortalitas pada peternakan ayam broiler dilakukan penelitian perbandingan metode *Naïve Bayes* dan C4.5. Hasil dari perbandingan akan dievaluasi menggunakan *confusion matrix*. Hasil dari pengujian data menggunakan *confusion matrix* menghasilkan nilai akurasi dari metode C4.5 lebih besar dari pada metode *Naïve Bayes*. Nilai akurasi dari metode C4.5 adalah 93% dan nilai akurasi dari metode *Naïve Bayes* adalah 88.66%.

Kata kunci: mortalitas; ayam broiler; *naïve bayes*; c4.5; perbandingan, prediksi.

ABSTRACT

Broiler chickens are the fastest type of livestock to harvest. But in raising broiler chickens, there must be many problems faced, for example the death rate. To reduce losses, farmers should pay attention to what factors cause the death of the chicken. Several studies that examined broiler chickens used experimental methods and RAL. However, there is still no research on the mortality of broilers using computing. To find out which method is better for predicting mortality at broiler chicken farms, a comparative study of Naïve Bayes and C4.5 methods is conducted. The results of the comparison will be evaluated using the confusion matrix. The results of testing the data using the confusion matrix resulted in an accuracy value of the C4.5 method greater than the Naïve Bayes method. The accuracy value of the C4.5 method is 93% and the accuracy value of the Naïve Bayes method is 88.66%

Keywords: mortality; broiler chicken; *naïve bayes*; c4.5; comparison; prediction.

1. PENDAHULUAN

Ayam broiler adalah jenis ayam jantan dan betina muda yang berumur sekitar 6-8 minggu yang dipelihara secara intensif, guna memperoleh daging yang optimal untuk memenuhi gizi [1]. Dalam pemeliharaan ayam broiler pasti banyak juga masalah yang dihadapi, salah satu contohnya adalah mengalami kerugian akibat kematian ayam broiler. Karena itu peternak ayam broiler harus mengetahui secara detail setiap langkah dan komponen penentu keberhasilan ternak ayam broiler [1].

Beberapa berita di kota-kota besar mengabarkan tentang kenaikan harga ayam broiler. Faktor kenaikan harga ayam broiler tersebut adalah mahalnya biaya produksi dan juga kurangnya pasokan ayam broiler yang dijual dipasaran. Salah satu faktor kurangnya pasokan ayam broiler adalah tingkat kematian pada ayam broiler yang tinggi [2].

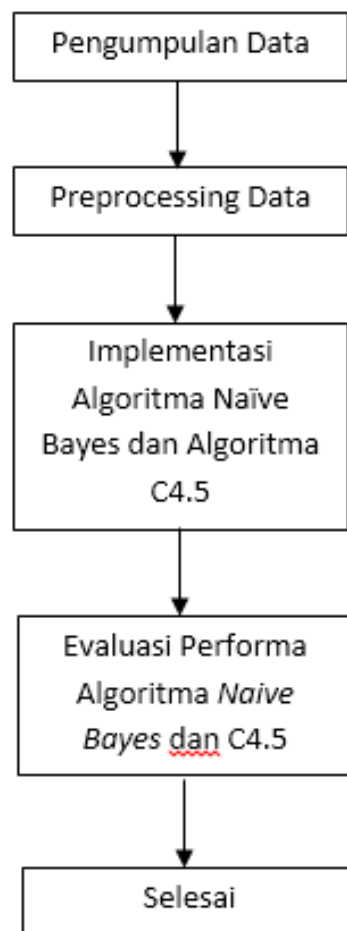
Berbagai penelitian tentang ayam broiler telah dilakukan namun belum ada yang menggunakan metode komputasi. Maka dari itu akan dilakukan penelitian tentang mortalitas ayam broiler.

Perbandingan Algoritma *Decision Tree* (C4.5) Dan *Naïve Bayes* Pada Data Mining Untuk Identifikasi Tumbuh Kembang Anak Balita menyimpulkan bahwa metode *Naïve Bayes* lebih baik dari metode C4.5 [3]. Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan antara Algoritma C4.5 dan *Naïve Bayes* menyimpulkan bahwa metode C4.5 lebih baik dari metode *Naïve Bayes* [4].

Dari latar belakang diatas tersebut akan dilakukan penelitian dengan membandingkan 2 metode yaitu metode *Naïve Bayes* dan metode C4.5 untuk melihat seberapa besar akurasi yang didapatkan.

2. METODOLOGI PENELITIAN

Berikut adalah kerangka penelitian yang akan dilakukan oleh peneliti yang akan dijelaskan pada gambar 1.



Gambar 1. Kerangka Penelitian

Berdasarkan Gambar 1. di atas, tahapan-tahapan dalam penelitian ini adalah sebagai berikut :

2.1 Pengumpulan Data

Tahap pertama pada penelitian ini adalah pengumpulan data dari sumber data yaitu peternakan ayam broiler. Pengumpulan data ini dilakukan dengan cara mengunjungi beberapa peternakan yang bekerja sama dengan PT. Semesta Mitra Sejahtera agar mendapatkan data yang valid.

Data yang didapatkan berjumlah 300 data kandang. Data yang digunakan adalah data pada tahun 2018 bulan Maret. Data yang digunakan juga data peternakan yang berada di Jawa Timur karena Jawa Timur adalah pemasok produksi ternak tingkat nasional [5].

Variabel yang didapatkan adalah jumlah populasi ayam broiler pada tiap kandang, jenis kandang yang digunakan, jenis ayam broiler yang ditanakkan, BW (*Body Weight*) atau bobot ayam broiler, umur panen ayam broiler dan yang terakhir mortalitas ayam broiler.

2.2 Preprocessing Data

Preprocessing Data adalah suatu proses/langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas [6]. Dalam *preprocessing data*, yang akan dilakukan adalah melakukan transformasi data.

Proses transformasi data tersebut dilakukan menggunakan pengklasteran pada tiap atribut. Pengklasteran tersebut ditetapkan dari peraturan PT. Semesta Mitra Sejahtera. Peraturan tersebut merupakan peraturan yang tidak tertulis yang didapatkan dengan melakukan wawancara terhadap *Branch Head* dari PT. Semesta Mitra Sejahtera. Berikut adalah pengklasteran dari tiap atribut.

- a) Populasi
 Populasi dibagi menjadi 3 kelas yaitu kecil, sedang dan besar. Nilai dari tiap kelas adalah :
 - 1) Kecil = 2500 – 7500
 - 2) Sedang = 8000 – 15000
 - 3) Besar = ≥ 16000
- b) Jenis kandang
 Jenis kandang dibagi menjadi 2 yaitu open house dan close house.
- c) Jenis ayam
 Jenis ayam dibagi menjadi 2 jenis yaitu COBB dan ROSS.
- d) BW(*Body Weight*)
 BW(*Body Weight*) dibagi menjadi 3 kelas yaitu :
 - 1) Kurang = < 1.7 kg
 - 2) Sedang = 1.7 – 2 kg
 - 3) Besar = > 2 kg
- e) Umur panen
 Umur panen dibagi menjadi 3 kelas yaitu :
 - 1) Kurang = < 35 hari
 - 2) Sedang = 35 hari
 - 3) Lebih = > 35 hari
- f) Mortalitas
 Mortalitas dibagi menjadi 2 kelas yaitu :
 - 1) Besar = > 4
 - 2) Kecil = ≥ 4

2.3 Implementasi Algoritma Naive Bayes Dan Algoritma C4.5

Pengujian data yang dilakukan untuk memprediksi mortalitas pada peternakan ayam broiler adalah menggunakan *K-Fold Cross Validation*. *K-Fold Cross Validation* berguna untuk membagi kumpulan data menjadi data latih dan data uji. *K-Fold Cross Validation* digunakan karena merupakan salah satu metode yang terbaik untuk memvalidasi data yang akan digunakan. Salah satu contoh dari *K-Fold Cross Validation* adalah *10-Fold Cross Validation*. Teknik ini akan membagi kumpulan data menjadi 10 subset dengan ukuran yang sama, sembilan dari 10 subset data digunakan untuk pelatihan, sementara satu subset yang tertinggal digunakan untuk pengujian. Proses diulang selama sepuluh kali, dan hasil akhirnya diperkirakan sebagai tingkat kesalahan rata-rata pada contoh uji.

2.4 Evaluasi Performa Algoritma Naive Bayes Dan Algoritma C4.5

Evaluasi performa algoritma *Naive Bayes* dan algoritma C4.5 menggunakan *confusion matrix* untuk menghitung Akurasi, Presisi dan *Recall*. Berikut adalah tabel dari *confusion matrix* :

Tabel 1. Tabel Confution Matrix [7]

		<i>Dectected</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Actual</i>	<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
	<i>Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Akurasi adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual. Presisi adalah tingkat ketepatan antara informasi yang diminta dengan jawaban yang diberikan oleh sistem. *Recall* adalah

tingkat keberhasilan sistem dalam menemukannya kembali informasi. Berikut perhitungannya Akurasi, Presisi, dan *Recall*. Berikut adalah rumus untuk perhitungan Akurasi, Presisi, dan *Recall* [8]:

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \quad (1)$$

$$Presisi = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil Implementasi Algoritma Naive Bayes dan Algoritma C4.5

Untuk menentukan berapa nilai K yang akan digunakan pada *K-Folds Cross Validation* dilakukan pengujian nilai K. Nilai K yang diuji adalah 5, 10 dan 15. Pengujian dilakukan pada aplikasi Weka 3.6. Perbandingan nilai dapat dilihat pada tabel 2.

Tabel 2. Perbandingan pengujian nilai K

	<i>Naive Bayes</i>			<i>C4.5</i>		
	5	10	15	5	10	15
<i>Correctly</i>	89%	88.667%	88.667%	93%	93%	93%
<i>Incorrectly</i>	11%	11.33%	11.33%	7%	7%	7%
<i>Relative absolute error</i>	65.8431%	63.9084%	64.2237%	58.237%	57.145%	57.3599%

Setelah dilakukan pengujian pada nilai K tersebut, dapat dilihat nilai 10 adalah nilai yang terbaik diantara nilai yang diuji. Maka dari hasil pengujian tersebut akan dipilih nilai 10 untuk digunakan dalam *K-Folds Cross Validation*.

Pada proses *10-Folds Cross Validation*, data yang berjumlah 300 akan dibagi rata menjadi 10 bagian yang mana 9 dari 10 data akan digunakan sebagai data latihan dan 1 data yang lain digunakan untuk data uji. Proses dilakukan berulang 10 kali sehingga tiap data akan dicoba sebagai data uji. Setelah itu hasil akan dirata-rata untuk mendapatkan nilai akhir. Pada tabel 3 adalah nilai dari setiap *10-Folds Cross Validation* yaitu 10 kali perhitungan menggunakan metode *Naive Bayes*.

Tabel 3. Tabel Perhitungan 10-Folds Cross Validation Metode Naive Bayes

<i>Fold</i>	<i>Metode Naive Bayes</i>		
	<i>Akurasi</i>	<i>Presisi</i>	<i>Recall</i>
1	90%	0%	0%
2	93.33%	0.66%	0.66%
3	90%	0.5%	0.33%
4	83.33%	0%	0%
5	90%	0%	0%
6	90%	0%	0%
7	76.66%	0%	0%
8	90%	0%	0%
9	90%	0%	0%
10	93.33%	0.21%	0.13%

Selanjutnya akan dilakukan perhitungan rata-rata pada nilai akurasi, nilai recall dan nilai presisi.

Tabel 4. Tabel perhitungan rata-rata 10-Folds cross validation metode naive bayes

<i>10- Fold Cross Validation</i>	
<i>Akurasi</i>	88.66%
<i>Presisi</i>	0.21%
<i>Recall</i>	0.13%

Nilai akurasi dari metode *Naïve Bayes* sudah baik namun bila dilihat nilai dari presisi dan nilai dari *recall* sangatlah kecil. Nilai tersebut kecil dikarenakan dari banyaknya angka nol yang didapat dari perhitungan nilai presisi dan nilai *recall* maka bila dirata-rata akan mendapat nilai yang relatif kecil. Pada tabel 5 adalah nilai dari setiap *10-Folds Cross Validation* yaitu 10 kali perhitungan menggunakan metode C4.5

Tabel 5. Tabel perhitungan 10-Folds cross validation metode C4.5

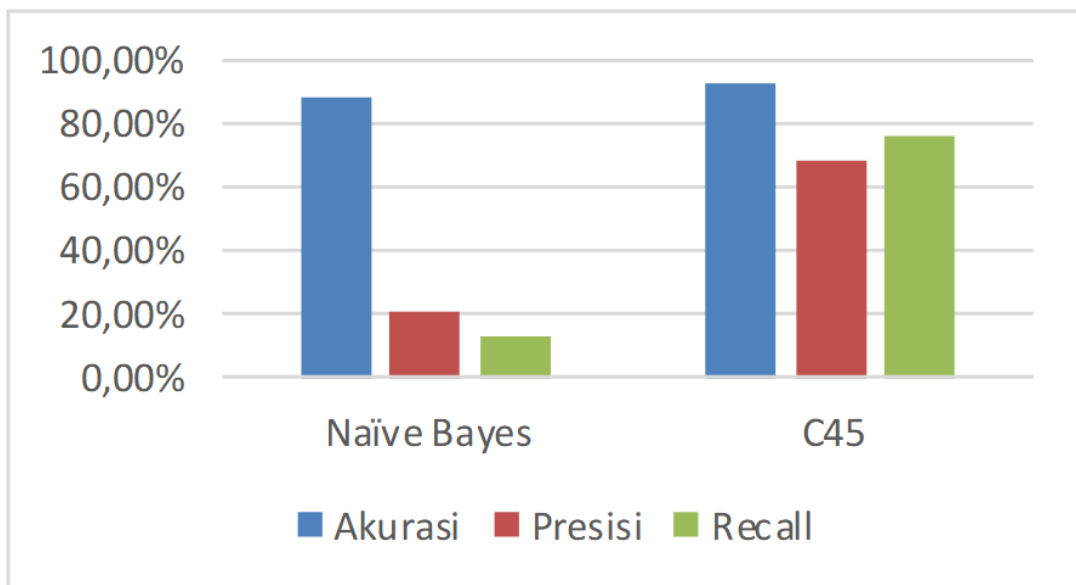
<i>Fold</i>	<i>Metode C4.5</i>		
	<i>Akurasi</i>	<i>Presisi</i>	<i>Recall</i>
1	93.33%	0.66%	0.66%
2	90%	0.5%	0.66%
3	93.33%	0.6%	1%
4	90%	0.5%	1%
5	96.66%	0.75%	1%
6	96.66%	1%	0.66%
7	80%	0.2%	0.33%
8	93.33%	0.66%	0.66%
9	100%	1%	1%
10	96.66%	1%	0.66%

Selanjutnya akan dilakukan perhitungan rata-rata pada nilai akurasi, nilai recall dan nilai presisi.

Tabel 6. Tabel perhitungan rata-rata 10-Folds cross validation metode C4.5

<i>10- Fold Cross Validation</i>	
<i>Akurasi</i>	93.66%
<i>Presisi</i>	0.68%
<i>Recall</i>	0.76%

Nilai akurasi dari metode C4.5 sudah baik dan nilai dari presisi dan nilai *recall* juga bisa dikatakan besar. Nilai tersebut besar dikarenakan dari setiap perhitungan tidak ada yang bernilai nol, jadi bila dirata-rata akan mendapatkan nilai yang relatif besar. Bila digambarkan dalam bentuk grafik sebagai berikut seperti pada gambar 2 :



Gambar 2. Grafik Perbandingan Metode Naïve Bayes dan Metode C4.5

3.2 Pembahasan

Tabel 7. Tabel hasil perbandingan metode *naïve bayes* dan metode C4.5

<i>Evaluation</i>	<i>10- Fold Cross Validation</i>	
	<i>Naïve Bayes</i>	<i>C4.5</i>
<i>Accuration</i>	88.66%	93%
<i>Precision</i>	0.21%	0.68%
<i>Recall</i>	0.13%	0.76%

Tingkat akurasi dari metode C4.5 lebih tinggi daripada metode *Naïve Bayes* dikarenakan dalam metode C4.5 memprediksi kebenaran lebih banyak dibandingkan dengan metode *Naïve Bayes*. Tingkat presisi dari metode C4.5 lebih tinggi daripada metode *Naïve Bayes* dikarenakan tingkat ketepatan informasi yang diminta dengan jawaban yang diberikan oleh sistem lebih banyak.

Tingkat *recall* dari metode C4.5 lebih tinggi daripada metode *Naïve Bayes* dikarenakan tingkat keberhasilan sistem dalam menemukan kembali informasi lebih tinggi daripada menggunakan metode *Naïve Bayes*. Pada evaluasi menggunakan *confution matrix* dengan metode *Naïve Bayes* didapatkan data sebagai berikut :

Tabel 8. Tabel *confusion matrix* metode *naïve bayes*

		<i>Detected</i>	
		<i>Positive</i>	<i>Negative</i>
Actual	<i>Positive</i>	4	26
	<i>Negative</i>	8	262

Hasil dari pengujian mendapatkan 266 data dari total 300 data diklasifikasikan dengan benar dan 34 data dari total 300 data dikatakan salah. Jika dilihat kembali maka akan terlihat berapa jumlah mortalitas besar yang benar dan salah maupun mortalitas kecil yang benar atau salah.

Dari tabel 8 diketahui nilai dari TP (*True Positive*) adalah 4 data karena terdapat 4 data adalah mortalitas besar yang diklasifikasikan benar. Nilai dari TN (*True Negative*) adalah 262 data karena terdapat 262 data mortalitas kecil yang diklasifikasikan benar.

Nilai dari FP (*False Positive*) adalah 8 data karena terdapat 8 data mortalitas kecil yang diklasifikasikan salah. Nilai dari FN (*False Negative*) adalah 26 data karena terdapat 26 data mortalitas besar diklasifikasikan yang salah.

Pada evaluasi menggunakan *confution matrix* dengan metode C4.5 didapatkan data sebagai berikut :

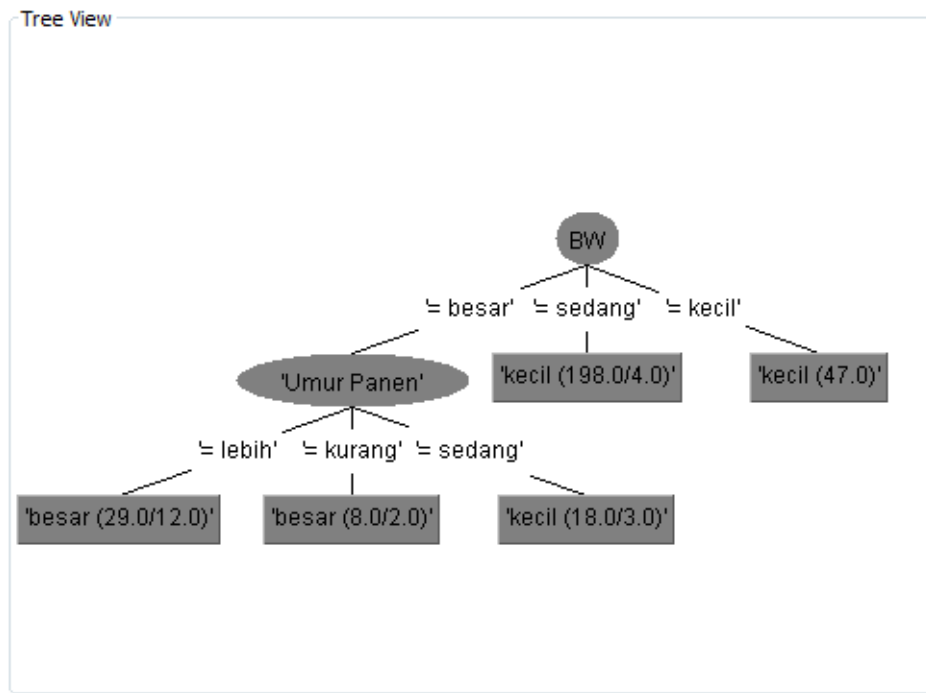
Tabel 9. Tabel *confusion matrix* metode C4.5

		<i>Detected</i>	
		<i>Positive</i>	<i>Negative</i>
Actual	<i>Positive</i>	23	7
	<i>Negative</i>	14	256

Hasil dari pengujian mendapatkan 279 data dari total 300 data diklasifikasikan dengan benar dan 21 data dikatakan salah. Jika dilihat kembali maka akan terlihat berapa jumlah mortalitas besar yang benar dan salah maupun mortalitas kecil yang benar atau salah.

Dari tabel 4.9 diketahui nilai dari TP (*True Positive*) adalah 23 data karena terdapat 23 data adalah mortalitas besar yang diklasifikasikan benar. Nilai dari TN (*True Negative*) adalah 256 data karena terdapat 256 data mortalitas kecil yang diklasifikasikan benar. Nilai dari FP (*False Positive*) adalah 14 data karena terdapat 14 data mortalitas kecil yang diklasifikasikan salah. Nilai dari FN (*False Negative*) adalah 7 data karena terdapat 7 data mortalitas besar diklasifikasikan yang salah.

Dari pemrosesan yang telah dilakukan menggunakan metode C4.5 dengan *10-Fold Cross Validation* menghasilkan tree sebagai berikut :



Gambar 3. Gambar Tree Keputusan Metode C4.5

Dari gambar 4.3 dapat dilihat *Body Weight* adalah node pertama dari tree keputusan. *Tree* tersebut memiliki 5 rule dan 6 cabang. *Tree* tersebut hanya menampilkan 2 dari 5 variabel yang ada karena ternyata hanya 2 variabel tersebut yang sangat mempengaruhi mortalitas pada ayam broiler.

Dari hasil penelitian yang telah dilakukan dan penjelasan dari perbandingan antara metode *Naïve Bayes* dan metode C4.5 dapat disimpulkan bahwa untuk penelitian perbandingan metode *Naïve Bayes* dan metode C4.5 untuk memprediksi mortalitas pada peternakan ayam broiler adalah metode C4.5 yang terbaik dikarenakan metode C4.5 menghasilkan akurasi, presisi dan *recall* yang lebih besar dari pada metode *Naïve Bayes*.

4. KESIMPULAN

Berikut adalah kesimpulan yang dapat diambil, yaitu :

- Penelitian perbandingan metode *Naïve Bayes* dan metode C4.5 untuk memprediksi mortalitas pada peternakan ayam broiler adalah metode C4.5 yang terbaik dikarenakan metode C4.5 menghasilkan akurasi sebesar 93%, presisi sebesar 0.68% dan *recall* sebesar 0.76%.
- Tree* keputusan yang dihasilkan dari metode C4.5 menampilkan dua variabel yaitu variabel *Body Weight* dan variabel umur panen. Variable tersebut adalah variabel yang paling berpengaruh untuk memprediksi mortalitas pada ayam broiler.

DAFTAR PUSTAKA

- Nuril, B., Ubaidillah, M. (2013). Pengaruh Frekuensi Penyemprotan Desinfektan Pada Kandang Terhadap Jumlah Kematian Ayam Broiler.
- Republika.co.id. (2017). Harga Ayam Potong Naik di Malang
- Listiana., Mila. (2015). Perbandingan Algoritma Decision Tree (C4.5) Dan *Naïve Bayes* Pada Data Mining Untuk Identifikasi Tumbuh Kembang Anak Balita.
- Iskandar, D., Suprpto, Y. (2013). Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan antara Algoritma C4.5 dan *Naïve Bayes*.
- Tribunnews. (2016). Jawa Timur Jadi Andalan Peternakan Nasional.
- Mayadewi, P., Roseli, E. (2015). Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining.
- Susanto, Heri. (2014). Data Mining Untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan Dan Prestasi Masa Lalu.

- [8] Tampubolon, K., Saragih, H., Reza, B. (2013). Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan.