

Implementasi Algoritma C4.5 Untuk Klasifikasi Produktivitas Bebek Petelur Peternakan Tradisional

Willy Rumapea^{1*}, Ummul Khair²

^{1,2}Teknik Informatika, Fakultas Teknik dan Komputer, Universitas Harapan Medan, Indonesia

^{1*}willyrumapea123jr@gmail.com, ²ummul.kh@gmail.com

^{*)} willyrumapea123jr@gmail.com

Abstrak-Pemanfaatan teknologi dalam berbagai bidang dapat dimanfaatkan untuk menyelesaikan banyak masalah terkait. Salah satunya dalam bidang peternakan. Masalah penentuan produktivitas bebek dalam peternakan bebek Ibu Juntak merupakan salah satu yang paling sering ditemui. Pengambilan keputusan apakah seekor bebek termasuk kategori produktif atau tidak merupakan tindakan krusial yang menentukan keuntungan yang didapat dari hasil penjualan telur bebek tersebut. Di lain sisi permintaan pasar yang tinggi menuntut peternak seperti Ibu Juntak untuk selalu memberikan produk dengan jumlah cukup dan kualitas yang tinggi. Maka dari itu penentuan bebek yang punya produktivitas baik sangat diperlukan untuk mendukung permintaan dan penjualan. Dalam penelitian ini, dilakukan klasifikasi pemrosesan data menggunakan Algoritma C4.5 yang merupakan salah satu metode dalam membangun model pohon keputusan. Algoritma C4.5 sendiri merupakan pengembangan dari metode sebelumnya yaitu ID3 yang lebih sensitif dalam menentukan keputusan dari fitur yang digunakan. Tujuan dari penelitian ini adalah mendapatkan faktor penentu beserta turunannya yang paling berpengaruh terhadap produktivitas bebek petelur. Hasil akhir dari penelitian adalah pohon keputusan bercabang yang mengandung informasi tentang faktor penentu produktivitas seekor bebek petelur yang menunjukkan hasil bahwa pakan merupakan faktor paling utama dari produktivitas bebek petelur dibandingkan dengan faktor lain yang digunakan.

Kata Kunci: Produktivitas, C4.5, Pohon Keputusan, Klasifikasi, Bebek

Abstract-The utilization of technology across various sectors has proven effective in addressing numerous related challenges, including in the field of animal husbandry. One recurring issue in the duck farming business owned by Mrs. Juntak is the determination of duck productivity. Making decisions on whether a duck is categorized as productive or not is a critical process that directly impacts the profit gained from the sale of its eggs. On the other hand, high market demand compels farmers like Mrs. Juntak to consistently provide products in sufficient quantity and with high quality. Therefore, identifying ducks with high productivity is essential to support both market demand and sales performance. This study applies data classification using the C4.5 algorithm, a well-established method for constructing decision tree models. The C4.5 algorithm is an extension of the earlier ID3 method, offering greater sensitivity in determining decisions based on the features used. The aim of this research is to identify the most influential determining factors and their derivatives that affect the productivity of laying ducks. The final output of the study is a branched decision tree that contains valuable insights into the key factors influencing duck productivity. The results indicate that feed is the most significant factor affecting the productivity of laying ducks, compared to other variables considered in the study.

Keywords: Productivity, C4.5, Decision Tree, Classification, Duck

1. PENDAHULUAN

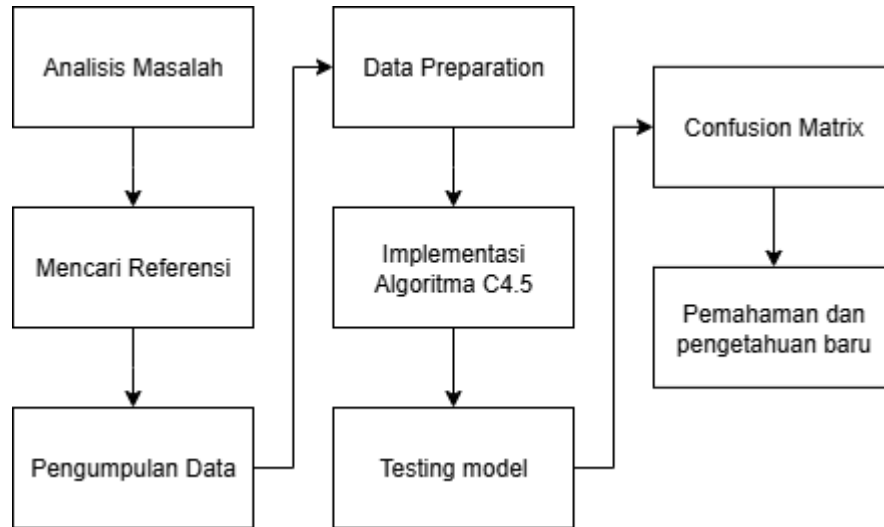
Telur bebek merupakan salah satu sumber pangan dan protein hewani yang cukup digemari oleh masyarakat luas. Rasa yang enak menjadi salah satu alasan mengapa telur bebek cukup menjadi primadona sebagai asupan protein dan lauk di Indonesia serta harganya yang terjangkau. Kandungan gizi dalam telur bebek juga lebih tinggi jika dibandingkan dengan telur dari unggas lain. Telur bebek mengandung asam amino esensial, asam omega-3, albumin dan beberapa zat lain yang dibutuhkan tubuh dalam jumlah yang lebih banyak [1]. Peternakan bebek petelur dapat dijadikan sebagai sumber penghasilan yang cukup menjanjikan dan menguntungkan, bahkan menurut Badan Pusat Statistika pada tahun 2020 produksi telur bebek mencapai angka 40.104 ton [2]. Namun pada kenyataannya banyak masalah yang menjadi tantangan beresiko cukup tinggi menggagalkan usaha ternak ini, salah satunya adalah masalah produktivitas bebek itu sendiri. Hal ini bisa terjadi karena bagi masyarakat peternakan masih identik dengan tata cara tradisional [3]. Masalah produktivitas ini merupakan masalah dimana bebek yang

sudah cukup umur untuk memasuki masa produktivitas bertelur tidak mampu menghasilkan telur dengan jumlah yang maksimal selama satu tahun. Dengan menggunakan sistem peternakan tradisional maka rata rata produksi telur selama setahun sekitar 100-150 butir. Masalah ini cukup vital karena biaya untuk pakan ternak mencapai 70% dari total modal [4]. Bisa diartikan pemberian pakan tidak tepat kepada bebek yang memang punya produktivitas dalam bertelur. Sedangkan bebek yang tidak produktif perlu dipisahkan sembari mengusahakan bebek tersebut kembali produktif. Diperlukan sebuah cara yang digunakan untuk menyelesaikan permasalahan tersebut. Menemukan faktor yang menjadi masalah utama dalam produktivitas bebek petelur langkah yang paling tepat untuk menyelesaikan masalah tersebut. Untuk mendapatkan faktor-faktor tersebut secara akurat dibutuhkan perhitungan secara tepat dan sistematis, sehingga hasilnya dapat secara tepat menyelesaikan masalah produktivitas ini. Teknologi komputer saat ini sedang banyak dimanfaatkan adalah *machine learning*. *Machine learning* merupakan salah satu cabang keilmuan kecerdasan buatan atau Artificial Intelligence (AI) [5]. *Machine learning* merupakan sebuah program komputer yang dirancang untuk memberikannya kemampuan untuk belajar sendiri dari data yang tersedia tanpa harus diprogram ulang secara eksplisit [6]. Tujuan dari *machine learning* adalah untuk mendapatkan informasi melalui berbagai proses dari kumpulan data besar atau dataset [7]. *Machine learning* memiliki banyak algoritma yang dapat diimplementasikan sesuai dengan jenis masalah yang akan dipecahkan. *Machine learning* memiliki cara belajar yang beragam salah satunya adalah *supervised learning* yang merupakan teknik belajar yang dibimbing. Dibimbing yang dimaksud adalah dengan mengarahkan model *machine learning* tersebut untuk mempelajari dan mengklasifikasikan data berdasarkan tanda atau label yang sudah diberikan sebelumnya pada data yang akan digunakan [8]. Salah satu algoritma *machine learning* dalam *supervised learning* adalah *decision tree*. *Decision tree* adalah satu satu algoritma dalam masalah klasifikasi yang paling populer dan cukup mudah dipahami [9]. *Decision tree* adalah algoritma yang digunakan untuk membuat pohon keputusan sebagai prediktor hasil atau menjadi acuan dalam pengambilan keputusan dari data yang tersedia. *Decision tree* juga memiliki beberapa metode dalam pembuatan dan pemodelannya, salah satunya adalah C4.5 yang merupakan hasil pengembangan metode sebelumnya ID3. Algoritma C4.5 akan membangun dan memodelkan pohon keputusan dari *root node* yang merupakan atribut paling besar pengaruhnya dalam proses klasifikasi. Selanjutnya cabang akan terbentuk dari setiap node atau simpul dan proses itu berulang hingga semua atribut digunakan atau semua nilai menjadi homogen dan berakhir dengan *leaf node* [10]. Algoritma C4.5 memprediksi dan memisahkan nilai atribut berdasarkan nilai akhir yang disebut *gain ratio*, dimana *gain ratio* adalah nilai yang menunjukkan seberapa informatif sebuah atribut dan pengaruhnya dalam pembentukan pohon keputusan. Perbedaan utama dari C4.5 dan ID3 adalah *gain ratio* sebagai patokan pemilihan atribut pada C4.5 sedangkan ID3 menggunakan *information gain*. Selain itu C4.5 akan memangkas node yang tidak perlu sehingga pohon keputusan yang dihasilkan lebih tepat dan akurat [11]. Pada sebuah penelitian dinyatakan untuk mencari atribut paling berpengaruh untuk produksi ikan air tawar, didapat atribut luas kolam sebagai atribut paling berpengaruh terhadap keberhasilan produksi ikan air tawar dari 19 atribut didalam dataset. Hasil keakuratan dari *Confusion Matrix* juga memberikan angka 94,44% menunjukkan modelnya mampu mengklasifikasi setiap data dengan baik [12]. Penelitian lainnya dalam memantau kesehatan domba dengan data penting seperti suhu tubuh dan detak jantung domba. Hasil *accuracy*, *recall*, dan *precision* yang didapat rata-rata diatas 80% menunjukkan sistem mampu menghitung prediksi kesehatan domba dengan sangat baik [13]. Tujuan penelitian ini untuk mencari faktor-faktor yang berpengaruh dalam produktivitas bebek petelur sehingga dapat dijadikan acuan bagi para peternak bebek. Penelitian ini dapat langsung memberikan manfaat karena hasil dapat diterapkan sedini mungkin untuk mencegah produktivitas bebek yang rendah. Melalui pendekatan ini hasil penelitian ini tidak hanya berkontribusi dalam pada teknologi pengembangan AI tetapi juga mendukung para pelaku usaha terkhusus dalam bidang peternakan untuk mengembangkan usahanya dan matematis dalam menyelesaikan masalah produktivitas bebek petelur.

2. METODOLOGI PENELITIAN

Pada penelitian ini terdapat beberapa tahapan yang tersusun secara terstruktur untuk mendapatkan hasil maksimal dari pengembangan model algoritma C4.5 untuk mendapatkan faktor-faktor pendukung produktivitas bebek petelur. Pada metode penelitian terdapat beberapa proses meliputi proses pengumpulan data, implementasi algoritma, dan evaluasi untuk menilai kinerja model. Dalam penelitian ini penulis melakukan riset terhadap

seorang peternak yaitu Ibu Juntak selaku pemilik peternakan bebek petelur Ibu Juntak. Tahapan utama dalam penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini merupakan data yang berisi tentang informasi umum mengenai peternakan bebek. Fitur atau atribut yang terkandung di dalamnya seperti pakan, jenis bebek, kandang, cuaca, bobot, bulu dan satu label yaitu produktivitas. Penulis mendapat dataset ini melalui wawancara kepada pemilik peternakan yaitu Ibu Juntak yang sebelumnya sudah memberikan izin riset dan menggunakan data yang ada dilapangan kepada penulis. Melalui dataset ini model akan dilatih dan ditesting sehingga dapat memisahkan bebek dengan pola atribut yang produktif. Berikut penjelasan singkat dari seluruh atribut yang akan digunakan: **Pakan:** jenis pakan yang diberikan kepada bebek yaitu pabrik dan campuran, **Jenis Bebek:** jenis bebek yang terdapat pada dataset ada 4 jenis yaitu bebek Medan, bebek Mojosari, bebek Hibrida, dan bebek Ratu MA, **Kandang:** adalah kondisi dalam kandang yaitu kering dan becek, **Cuaca:** jenis kondisi cuaca awalnya ukuran suhu lalu diubah kedalam bentuk kategorial yaitu sedang, terlalu panas, terlalu dingin, **Bobot:** berat badan bebek awalnya berbentuk kategorial lalu akan diubah kedalam angka pada proses *Data Preparation* berdasarkan rata rata berat bebek yaitu ideal, terlalu berat dan terlalu ringan, **Bulu:** adalah keadaan bulu pada tubuh bebek, jika bulu rontok potensi tidak bertelurnya tinggi, nilainya adalah rapat dan jarang. Lalu untuk label yaitu **Produktivitas:** tanda yang diberikan untuk model karena C4.5 adalah *Supervised Learning*, nilainya yaitu produktif dan tidak produktif.

Tabel 1. Sampel Dataset Produktivitas Bebek Petelur

Pakan	Jenis Bebek	Kandang	Cuaca	Bobot	Bulu	Produktivitas
Pabrik	Medan	Kering	Sedang	1.55	Rapat	Produktif
Pabrik	Medan	Kering	Sedang	1.68	Rapat	Produktif
Campuran	Mojosari	Becek	Terlalu dingin	1.12	Rontok	Tidak produktif
Campuran	Mojosari	Becek	Terlalu dingin	1.29	Rontok	Tidak produktif
Pabrik	Hibrida	Kering	Sedang	1.53	Rapat	Produktif
Pabrik	Hibrida	Kering	Terlalu panas	1.80	Rapat	Produktif
Pabrik	Hibrida	Kering	Terlalu panas	1.76	Rapat	Produktif
Parbik	Ratu MA	Kering	Sedang	1.69	Rapat	Produktif

Campurann	Ratu MA	Becek	Terlalu dingin	1.24	Rontok	Tidak produktif
Campuran	Ratu MA	Becek	Terlalu dingin	1.28	Rontok	Tidak produktif

Pada Tabel 1 berisi sampel dari dataset yang akan digunakan dalam melatih model dengan algoritma C4.5. Jumlah sebenarnya dari total data yang akan dipakai adalah 300 *record* atau baris data. Namun sebelum bisa digunakan terlebih dahulu dataset harus melewati proses *Data Preparation*. Proses tersebut dilakukan untuk memastikan data siap digunakan pada saat implementasi algoritma C4.5.

2.2 Data Preparation

Pada proses ini data pada dataset akan dipersiapkan untuk diolah dan dilakukan perhitungan di tahap implementasi algoritma. Proses ini sangat penting dan krusial karena menyumbang sangat banyak persentase keberhasilan sebuah model akan berhasil memisahkan dan mengklasifikasikan data dengan akurat sehingga menghasilkan *output* yang tepat sasaran. *Data Preparation* terdiri dari beberapa tahapan namun dalam penelitian ini penulis hanya menggunakan dua yaitu *Data Cleaning* dan *Data Tranformation*. *Data Cleaning* merupakan proses membersihkan data dari kecacatan seperti *typo*, duplikat, nilainya aneh atau anomali dan nilainya kosong. Lalu *Data Transformation* dilakukan untuk menyamakan tipe data pada dataset menjadi satu yang sama dalam hal ini menjadi numerik [14]. Pada tahap *Data Cleaning* dilakukan manual dengan cek data secara satu per satu sedangkan pada tahap *Data Transformation* menggunakan *One-Hot Encoding* dan *Labeling Encoding* [15]. Setelah melewati proses tersebut didapat hasil dataset baru dengan pertambahan jumlah kolom hasil dari proses *Data Transformation* namun tidak mengubah informasi, hanya mentranformasikan tipe nilai yang berbeda menjadi saam agar nantinya model dapat bekerja lebih baik.

Tabel 2. Sampel Dataset Proses Data Preparation

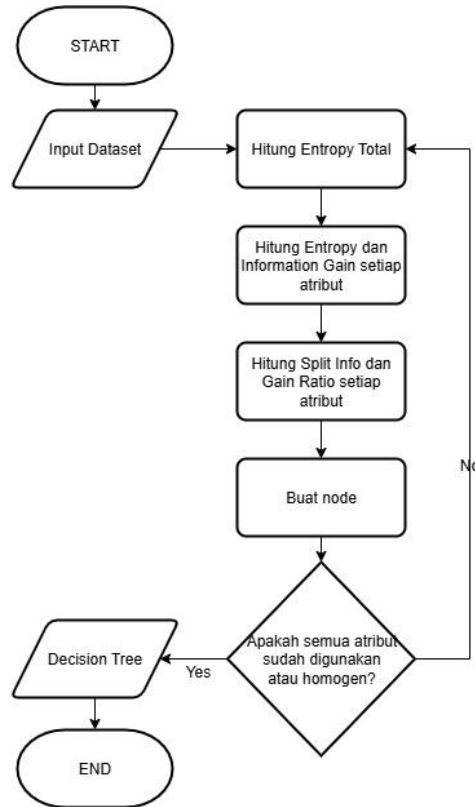
Pakan	Kandang	Cuaca	Bobot	Bulu	Produktivitas	Jenis bebek hibrida	Jenis bebek medan	Jenis bebek mojosari	Jenis bebek ratu ma
1	1	2	1.55	1	1	0	1	0	0
1	1	2	1.68	1	1	0	1	0	0
0	0	0	1.12	0	0	0	0	1	0
0	0	0	1.29	0	0	0	0	1	0
1	1	2	1.53	1	1	0	0	0	0
1	1	1	1.80	1	1	1	0	0	0
1	1	1	1.76	1	1	1	0	0	0
1	1	2	1.69	1	1	0	0	0	1
0	0	0	1.24	0	0	0	0	0	1
0	0	0	1.28	0	1	0	0	0	1

Tabel 2 diatas merupakan dataset dari sampel sebelumnya yang sudah melewati tahap *Data Preparation* sehingga terjadi perubahan pada tampilan datasetnya. Data sebelumnya memiliki beberapa *typo* atau salah penulisan, duplikasi dan nilai anomali sudah dibersihkan. Data sebelumnya memiliki tipe data campuran berupa teks atau bersifat kategorial diubah menjadi numerik. Proses tersebut dijalankan melalui tahap *Data Transformation* pada dengan metode *One-Hot Encoding* di atribut jenis bebek. Setiap jenis memiliki kolomnya sendiri dan jika jenis bebek benar pada baris tersebut diwakilkan dengan angka 1 dan jika tidak menjadi angka 0. Sedangkan *Labeling Encoding* dilakukan untuk atribut lain karena berdasarkan penuturan dari peternak yaitu Ibu Juntak ada beberapa nilai dalam atribut yang lebih unggul ketimbang nilai lain sehingga perlu diberikan bobot lebih agar model tau atribut yang lebih memiliki bobot lebih besar dalam pengaruhnya di dalam kasus produktivitas bebek petelur. Setelah proses ini selesai maka dataset sudah dapat dilakukan implementasi algoritma C4.5 dan dilakukan *training* serta *testing*.

2.3 Implementasi Algoritma C4.5

Proses ini merupakan inti dari penelitian ini dimana setiap langkah dan konsep dari Algoritma C4.5 akan diterapkan terhadap data sehingga akan memodelkan data tersebut menjadi bentuk *Decision Tree*. Setiap langkah

dari Algoritma C4.5 akan dilakukan perhitungan untuk setiap atribut. Untuk alur dan langkah-langkah dari algoritma C4.5 dapat dilihat pada Gambar 2.



Gambar 2. Flowchart Algoritma C4.5

Gambar 2 merupakan *flowchart* yang berisi alur algoritma C4.5 yang berjumlah 7 langkah. Gambar 2 adalah penjelasan dari algoritma C4.5 dapat dijabarkan dengan runut sebagai berikut:

- Input dataset yang sudah melewati proses *Data Preparation*.
- Hitung *entropy* dari setiap atribut.

Entropy dalam penelitian ini merupakan nilai yang menunjukkan ketidakpastian dalam dataset. Nilai ini menunjukkan seberapa beranekaragamnya data itu sendiri. *Entropy* digunakan untuk mengevaluasi impurity(ketidakmurnian) suatu dataset berdasarkan distribusi label [16]. Secara sistematis *entropy* dapat dihitung menggunakan rumus berikut:

$$E(S) = - \sum_{i=1}^n p_i * \log_2(p_i) \quad (1)$$

Keterangan:

S : himpunan kasus

n : jumlah partisi S

Pi : proporsi Si terhadap S

- Hitung *Information Gain*

Information Gain adalah nilai yang menunjukkan efektivitas setiap atribut dalam mengklasifikasikan data. *Information Gain* akan mengurangi nilai *entropy* total dengan nilai proporsi atribut yang sudah dikali dengan *entropy* atribut itu sendiri [17]. Rumus untuk mendapatkan *Information Gain* adalah sebagai berikut:

$$Gain(S, A) = E(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \cdot E(S_i) \quad (2)$$

Keterangan:

S : Jumlah kasus

A : Atribut

|S| : Jumlah kasus pada partisi ke-i

|S_i| : Jumlah kasus dalam S

d. Menghitung Split Info

Split Info adalah nilai ukuran seberapa besar persebaran data setelah dilakukan pembagian terhadap suatu atribut [18]. Tujuannya adalah untuk mengurangi nilai unik yang tidak berpengaruh terhadap pembentukan model *Decision Tree*. Sehingga hasil model lebih akurat dengan cabang atribut yang berkontribusi. Adapun rumus untuk menghitung *Split Info* adalah sebagai berikut:

$$Split Info(A) = \sum_{i=1}^n - \frac{|S_i|}{|S|} \cdot \log_2 \frac{|S_i|}{|S|} \quad (3)$$

Keterangan:

A : atribut dalam dataset

|S| : total kasus pada data

|S_i| : jumlah nilai subset terhadap total kasus atribut

e. Menghitung Gain Ratio

Gain Ratio adalah ukuran yang menunjukkan seberapa bermanfaat suatu atribut dalam pemisahan data berdasarkan klasifikasi dengan pembagian yang adil [19]. *Gain Ratio* lebih baik ketimbang *Information Gain* karena menggunakan nilai *Split Info* yang sudah membuang nilai yang tidak diperlukan dalam pembentukan model *Decision Tree*. Rumus dalam menghitung *Gain Ratio* adalah sebagai berikut:

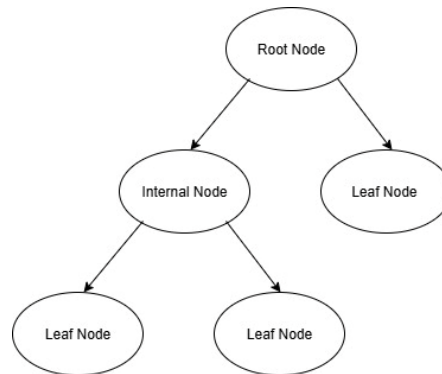
$$Gain Ratio(A) = \frac{Gain(A)}{Split Info(A)} \quad (4)$$

Keterangan:

A : atribut pada dataset

f. Penentuan Root Node dan cabang

Setelah semua atribut mendapatkan nilai *Gain Ratio* masing-masing selanjutnya dibandingkan setiap nilai untuk mendapatkan atribut dengan nilai *Gain Ratio* terbesar. Maka atribut tersebut akan menjadi *Root Node* dan memiliki pengaruh paling besar dalam pembentukan *Decision Tree* dalam hal ini faktor produktivitas bebek petelur. Setelah itu proses diulang untuk semua atribut yang tersisa hingga semua atribut digunakan atau semua data sudah homogen. Node paling akhir disebut *Leaf Node* yang berarti akhir dari cabang.



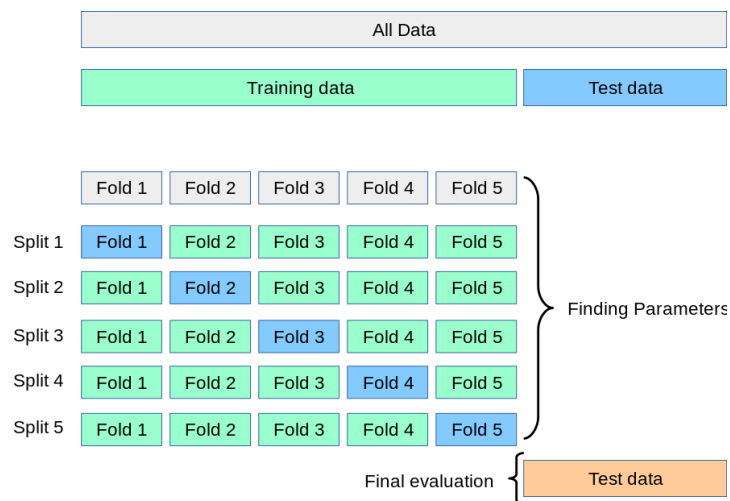
Gambar 3. Contoh Pembentukan dan Pemodelan Decision Tree

3. HASIL DAN PEMBAHASAN

Hasil penelitian ini adalah diagram *Decision Tree* serta evaluasi kinerja dan klasifikasi dari model dalam mengklasifikasikan faktor-faktor produktivitas bebek petelur berdasarkan 6 atribut prediktor ke dalam 2 kelas yaitu produktif dan tidak produktif. Model akan dilatih menggunakan data yang sudah melewati tahap *Data Preparation* dan dengan metode *testing* model *K-Fold Cross Validation*. Selain itu hasil klasifikasi model perlu untuk di evaluasi untuk melihat seberapa baik hasil klasifikasi dan pemisahan dari model itu sendiri. Sehingga dalam penelitian ini penulis menggunakan *Confusion Matrix* untuk mendapatkan nilai yang menjadi evaluasi dari hasil klasifikasi model.

3.1 Metode K-Fold Cross Validataion

Seperti yang sudah dijelaskan sebelumnya *K-Fold Cross Validation* adalah salah metode dalam pengujian model. Dalam metode jumlah dari data akan dibagi ke dalam beberapa *fold* tergantung banyaknya k [20]. Pada penelitian kali ini penulis menggunakan 10 k karena secara *default* k sebanyak 10 adalah yang paling ideal. Dari total 288 baris data yang sudah bersih akan dibagi kedalam 10fold. Karena tidak genap, secara default model akan membagi data kedalam beberapa fold 29 data dan sisanya 28 data agar data terbagi lebih merata. Pengujian akan dilakukan dengan menempatkan 1 data sebagai *testing* data dan sisanya sebagai *training* data. Sehingga pengujian akan adil karena semua data pernah mendapatkan peran yang sama. Ini juga berdampak ke dalam performa model, karena proses validasi yang dilakukan secara menyeluruh sehingga hasil lebih objektif dan stabil dalam mengklasifikasikan data baru.

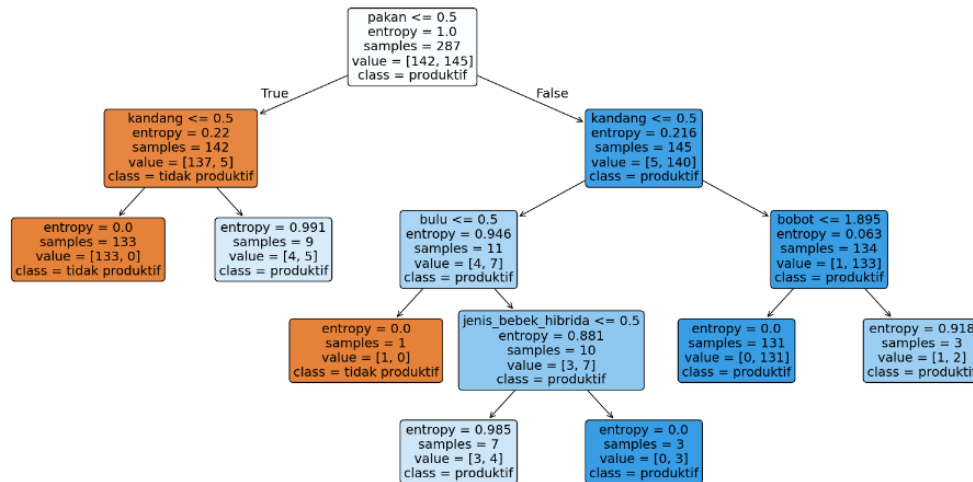


Gambar 4. Anologi Cara Kerja *K-Fold Cross Validation*

3.2 Hasil Decision Tree

Output dari model berupa grafik yang memvisualisasikan bentuk dari *Decision Tree* itu sendiri. Grafik akan berisi percabangan yang dimulai dari *Root Node* yang punya cabang hingga *Leaf Node*. Hasil ini yang dibutuhkan untuk menyelesaikan masalah produktivitas. Karena grafik akan menunjukkan semua atribut yang berpengaruh bersama dengan peneringkataannya berdasarkan tingkat pengaruhnya.

Pada gambar 5 ditampilkan hasil pemodelan dari *Decision Tree* dengan dua cabang utama dimana cabang sebelah kiri merupakan *true* dari persamaan dalam *node* atau simpul. Dimulai dari *Root Node* yang diisi oleh atribut pakan. Ini dapat diartikan bahwa faktor paling krusial dalam produktivitas bebek petelur berasal dari kualitas dan jenis pakan yang diberikan. Selanjutnya masuk ke cabangnya, dimana kedua cabang berisi atribut yang sama yaitu kandang.



Gambar 5. Grafik Visualisasi Decision Tree

Namun cabang sebelah kiri memiliki *class* tidak produktif. Itu berarti pakan dengan jenis campuran masuk ke *class* tidak produktif, juga dengan pola pakan campuran dan keadaan kandang kering berakhir dengan hasil tidak produktif. Untuk cabang sebelah kanan memiliki cabang yang lebih banyak serta lebih banyak atribut seperti bobot, bulu dan jenis bebek namun dengan jenis pakan yang berbeda yaitu jenis pakan pabrikan. Dengan pola pakan pabrik, kandang, bulu rapat dan jenis bebek hibrida akan berakhir dengan *LeafNode* yang memiliki *class* produktif. Namun jika bulu jarang masuk ke *class* tidak produktif. Begitu juga dengan pola pakan pabrik dan bobot ideal maka masuk ke dalam *class* produktif.

3.3 Evaluasi Confusion Matrix

Setelah mendapatkan hasil dari model berupa grafik pohon keputusan (Decision Tree), untuk memastikan hasil benar benar akurat menyelesaikan masalah klasifikasi perlu dilakukan evaluasi mendalam terhadap perhitungan yang didapat dari model. Dalam hal ini penulis menggunakan *Confussion Matrix* yang sudah disebutkan sebelumnya. Confusion Matrix merupakan prediksi yang menunjukkan jumlah prediksi model berdasarkan kategori aktual dan hasil prediksi [21]. Dalam kasus ini tabel akan berukuran 2 x 2 yang terdiri dari 4 komponen utama yaitu:

1. True Positive (TP): prediksi benar dan memang benar
2. True Negatif (TF): prediksi salah dan memang salah
3. False Positive (FP): prediksi benar padahal salah
4. False Negative (FN): prediksi salah padahal benar

Dari keempat nilai tersebut kemudian diambil untuk menentukan *accuracy*, *recall*, *precision* dan *F-1 Score* sebagai penilaian sejauh mana peforma dari model menghasilkan hasil klasifikasi secara keseluruhan.

🔍 Confusion Matrix:

```
[[140  5]
 [ 8 134]]
```

📄 Classification Report:

	precision	recall	f1-score	support
1	0.95	0.97	0.96	145
0	0.96	0.94	0.95	142
accuracy			0.95	287
macro avg	0.95	0.95	0.95	287
weighted avg	0.95	0.95	0.95	287

Gambar 6. Confusion Matrix



Berdasarkan gambar 6 model berhasil mengklasifikasikan total 145 kasus dengan label produktif yang diwakilkan dengan angka 1. Lebih spesifik lagi 140 TP dan hanya 5 FP. Menunjukkan model mampu mengurai lebih banyak kasus positive yang benar benar positive. Begitupun dengan label tidak produktif, model mengklasifikasikan 134 TN dan 8 FN. Model mampu membedakan kasus yang benar benar tidak produktif dengan selisih yang sangat jauh dengan kasus yang salah pisahkan.

3.1.1 Accuracy

Accuracy merupakan nilai yang mengukur tingkat klasifikasi yang tepat dari hasil perhitungan model. Pada gambar 6 tertulis angka 0.95 atau bisa dikatakan 95%, ini menunjukkan model akurat dalam memisahkan data dengan pola yang sesuai untuk label.

3.1.2 Precision

Precision mengukur proporsi nilai positif yang benar. Dengan kata lain mencari nilai benar yang memang benar. Pada gambar 6 angka precision dari masing-masing class adalah 0.95 dan 0.96. Dapat dikatakan menjadi 95% dan 96%. Berdasarkan ini dapat disimpulkan model sangat ketat dalam memisahkan data sehingga berhasil menemukan data yang valid sesuai dengan kelasnya.

3.1.3 Recall

Recall adalah nilai yang menunjukkan seberapa banyak data dalam yang benar berdasarkan kelas yang mampu ditemukan model dari total semua kasus dalam kelas. Berdasarkan gambar 6 diperoleh nilai masing-masing 0.97 untuk kelas produktif dan 0.94. Menunjukkan model akurat mengklasifikasikan data berdasarkan kelas yang ditentukan [22].

4. KESIMPULAN

Pakan merupakan faktor utama yang memengaruhi hasil produksi bebek petelur. Hal ini terlihat dari posisinya sebagai *Root Node* dalam model *Decision Tree*. Jumlah pakan kurang dari atau sama dengan 0,5 (campuran), cabang kiri dalam model menunjukkan kandang dengan kelas tidak produktif, artinya kualitas pakan yang rendah berdampak besar terhadap hasil produksi. Faktor lain yang turut memengaruhi meliputi kondisi kandang, bobot, bulu, dan jenis bebek. Model kini mampu menentukan urutan faktor utama secara akurat. Kondisi ini cocok dengan keadaan di lapangan, dimana peternak sering memberi pakan secara sembarangan seperti sisa makanan rumah atau tumbuhan liar yang tidak tentu memiliki gizi cukup. Faktor-faktor lain seperti kondisi kandang, berat badan, dan bulu sering kali diabaikan meskipun sangat penting untuk menjaga hasil produksi, terutama saat cuaca buruk. Hasil Confusion Matrix menunjukkan model memiliki kinerja yang sangat baik, dengan akurasi 95%, presisi 95%, dan recall 97% untuk kelas produktif begitupun pada kelas tidak produktif dengan angka di atas 90%. Model C4.5 berhasil mengungkap faktor penting yang perlu diperhatikan oleh peternak.

REFERENSI

- [1] A. Darwanto, N. Addina, and S. F. Wulandari, "Pemanfaatan Daun Mangga untuk Menurunkan Off-Odor Telur Itik dan Pengaruhnya Terhadap Kandungan Protein dan Lemak," *J. Kesehat.*, vol. 11, no. 2, pp. 124–129, 2023, [Online]. Available: <https://doi.org/10.25047/j-kes.v11i2.270>
- [2] S. Andanawari, P. Hartati, and Suharti, "Analisis Pendapatan Usaha Ternak Itik Petelur (Studi Kasus di Desa Kedungsari dan Desa Trasan Kecamatan Bandongan Kabupaten Magelang) Analysis," *J. Pengemb. Penyul. Pertan.*, vol. 18, no. 31, pp. 18–24, 2021.
- [3] A. E. Nugraha *et al.*, "PENINGKATAN PRODUKTIVITAS DAN DAYA SAING PETERNAK BEBEK PETELUR MELALUI PENERAPAN TEKNOLOGI PENCACAH TANAMAN, PEMBERSIH TELUR DAN DIGITAL MARKETING," *JMM (Jurnal Masy. Mandiri)*, vol. 9, no. 1, pp. 1–2, 2025.
- [4] Y. S. Melinda and A. Alatas, "Analisis Pendapatan Peternak Itik Petelur Sistem Pemeliharaan Intensif Di Nagari Koto Baru, Kecamatan IV Nagari, Kabupaten Sijunjung," *J. Agribus. Community Empower.*, vol. 7, no. 1, pp. 25–35, 2024, doi: 10.32530/jace.v7i1.736.
- [5] P. P. P. A. N. . F. I. R.H. Zer, B. H. Hayadi, and A. R. Damanik, "Pendekatan Machine Learning Menggunakan Algoritma C4.5 Berbasis Pso Dalam Analisa Pemahaman Pemrograman Website," *J.*





- Inform. dan Tek. Elektro Terap.*, vol. 10, no. 3, 2022, doi: 10.23960/jitet.v10i3.2700.
- [6] D. Juliani, "Implementasi Machine Learning untuk Klasifikasi Penyakit Kanker Paru Menggunakan Metode Naïve Bayes dengan Tambahan Fitur Chatbot," *J. Ilmu Pengetah. dan Teknol.*, vol. 8, no. 2, pp. 12–17, 2024, doi: 10.31543/jii.v8i2.351.
- [7] M. N. Waruwu, Yulisman Zega, Ratna Natalia Mendrofah, and Yakini Niat Telaumbanua, "Implementasi Algoritma Machine Learning Untuk Deteksi Performa Akademik Mahasiswa," *Tek. Teknol. Inf. dan Multimed.*, vol. 5, no. 2, 2024, doi: 10.46764/teknimedia.v5i2.214.
- [8] R. S. Nurhalizah, R. Ardianto, and P. Purwono, "Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review," *J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, pp. 61–72, 2024, doi: 10.54082/jiki.168.
- [9] F. Febriansyah, Z. Asti Dwiyantri, and Diash Firdaus, "Deteksi Serangan Low Rate Ddos Pada Jaringan Tradisional Menggunakan Machine Learning Dengan Algoritma Decision Tree," *Cyber Secur. dan Forensik Digit.*, vol. 6, no. 1, pp. 6–11, 2023, doi: 10.14421/csecurity.2023.6.1.3951.
- [10] A. R. Putra and Program, "Penerapan Teknik Data Mining Terhadap Prediksi Masa Studi Mahasiswa," vol. 1, pp. 33–37, 2024.
- [11] P. Anggita Kristanti and S. Jatmiko, "Klasifikasi Pemilihan Produk Perbankan Dengan Algoritma Decision Tree Id3 Dan C4.5," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 4, pp. 5999–6005, 2025, doi: 10.36040/jati.v9i4.13907.
- [12] A. Novareza, R. Herdiana, and Iin, "ANALISIS DATA MINING PADA PRODUKSI IKAN AIR TAWAR DI KECAMATAN PANJALU MENGGUNAKAN METODE C4.5 Apip," vol. 8, no. 1, pp. 252–258, 2024.
- [13] F. Nugraha, A. E. Wijaya, R. Hermawan, D. I. Sobari, and P. K. Domba, "Pengembangan Sistem IoT untuk Pemantauan Kesehatan Domba Dengan," vol. 17, no. 2, pp. 80–95, 2024.
- [14] P. Martins, F. Cardoso, P. Váz, J. Silva, and M. Abbasi, "Performance and Scalability of Data Cleaning and Preprocessing Tools: A Benchmark on Large Real-World Datasets," *Data*, vol. 10, no. 5, pp. 1–22, 2025, doi: 10.3390/data10050068.
- [15] E. Manai, M. Mejri, and J. Fattahi, "Impact of Feature Encoding on Malware Classification Explainability," *15th Int. Conf. Electron. Comput. Artif. Intell. ECAI 2023 - Proc.*, 2023, doi: 10.1109/ECAI58194.2023.10193964.
- [16] D. Maheswari, R. Bulan Anggraini, S. Aulia, Y. Diah, and B. Oloan Lubis, "Implementasi Algoritma C4.5 Untuk Klasifikasi Dampak Pola Penggunaan Media Sosial Terhadap Kesejahteraan Emosional," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 2, pp. 3411–3417, 2025, doi: 10.36040/jati.v9i2.13011.
- [17] A. Simanjorang, A. M. . Pardede, and S. Syahputra, "Implementasi Algoritma C4.5 untuk memprediksi Penjualan Paket Internet," *KAKIFIKOM (Kumpulan Artik. Karya Ilm. Fak. Ilmu Komputer)*, vol. 06, no. 02, pp. 106–115, 2024, doi: 10.54367/kakifikom.v6i2.4093.
- [18] V. M and L. C. Reddy, "A Review on Data mining from Past to the Future," *Int. J. Comput. Appl.*, vol. 15, no. 7, pp. 19–22, Feb. 2011, doi: 10.5120/1961-2623.
- [19] G. S. Reddy and S. Chittineni, "Entropy based C4.5-SHO algorithm with information gain optimization in data mining," *PeerJ Comput. Sci.*, vol. 7, pp. 1–22, 2021, doi: 10.7717/PEERJ-CS.424.
- [20] A. Nugraha, I. Maulana, P. Purwantoro, A. Susilo Yuda Irawan, and A. Voutama, "Analisis Penerapan Algoritma C4.5 Dalam Penentuan Siswa Penerima Beasiswa Karawang Cerdas," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 5, pp. 10078–10086, 2024, doi: 10.36040/jati.v8i5.10739.
- [21] C. Maulana Sidiq, A. Faqih, and G. Dwilestari, "Algoritma Decision Tree C4.5 Digunakan Untuk Mengklasifikasikan Data Stroke," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1869–1874, 2024, doi: 10.36040/jati.v8i2.8388.
- [22] N. S. B. Nugrahani and A. Prapanca, "Implementasi Algoritma C5.0 Pada Klasifikasi Status Gizi Balita Di Kecamatan Ponorogo," vol. 06, pp. 1089–1098, 2025.