

DECISION TREE SIMPLIFICATION THROUGH FEATURE SELECTION APPROACH IN SELECTING FISH FEED SELLERS

Esmi Nur Fitri*¹, Sri Winarno², Fikri Budiman³, Asih Rohmani⁴, Junta Zeniarja⁵, Edi Sugiarto⁶

^{1,2,4}Jurusan Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Indonesia

^{3,6,7}Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Indonesia

Email: ¹112201906175@mhs.dinus.ac.id, ²sri.winarno@dsn.dinus.ac.id, ³fikri.budiman@dsn.dinus.ac.id,
⁴aseharsoyo@dsn.dinus.ac.id, ⁵junta@dsn.dinus.ac.id, ⁶edi.sugiarto@dsn.dinus.ac.id

(Naskah masuk: 12 Desember 2022, Revisi : 17 Desember 2022, diterbitkan: 23 Maret 2023)

Abstract

Feed is a crucial variable because it can determine the success of fish farming. Breeders can use two types of artificial feed, namely alternative feed and pellets. Many cultivators need pellets as the main consumption for the fish they are cultivating because the pellets contain a composition that has been adjusted to their needs based on the type and age of the fish. However, currently, cultivators are facing a problem, namely the high price of fish pellets on the market. Therefore, an analysis of the classification of the selection of fish feed sellers is needed that is adjusted to several criteria like the number of types of feed, price, order, delivery, and availability of discounts. This study conducted a classification analysis of simplification of characteristics in selecting fish feed sellers in Kendal Regency that would then be compared with a model without feature selection by utilizing the Decision Tree C4.5 method. The results of this study are the decision tree with the best performance where C4.5 with the application of the selected feature has an accuracy value of 92%, while C4.5 without the selection feature has an accuracy of 86.8%. The results of this study indicate that the C4.5 method with the application of selection features is better than C4.5 without selection features so that it can be applied to the selection of freshwater fish feed sellers in Kendal Regency.

Keywords: *Classification, Decision Tree, Feature Selection, Fish Feed.*

PENYEDERHANAAN POHON KEPUTUSAN MELALUI PENDEKATAN SELEKSI FITUR DALAM MEMILIH PENJUAL PAKAN IKAN

Abstrak

Pakan adalah salah satu variabel yang krusial karena dapat menentukan keberhasilan budidaya ikan yang dilakukan. Pembudidaya dapat menggunakan dua jenis pakan buatan yaitu pakan alternatif dan pelet. Banyak pembudidaya yang membutuhkan pelet sebagai konsumsi utama bagi ikan yang dibudidayakan karena pelet mengandung komposisi yang telah disesuaikan dengan kebutuhan berdasarkan jenis dan umur ikan. Namun, saat ini pembudidaya menghadapi permasalahan yang cukup meresahkan yaitu sulitnya pengambilan keputusan untuk memilih penjual pakan ikan yang tepat di pasaran. Oleh karena itu, diperlukan sebuah analisis klasifikasi pemilihan penjual pakan ikan yang disesuaikan dengan beberapa kriteria seperti jumlah jenis pakan, harga, pemesanan, pengiriman, dan ketersediaan diskon. Penelitian ini melakukan sebuah analisis klasifikasi penyederhanaan kriteria dalam pemilihan penjual pakan ikan di Kabupaten Kendal yang kemudian akan dibandingkan dengan model tanpa seleksi fitur dengan memanfaatkan metode *Decision Tree C4.5*. Hasil dari penelitian ini adalah pohon keputusan dengan performa terbaik dimana *C4.5* dengan penerapan fitur seleksi memiliki nilai akurasi yaitu 92%, sedangkan untuk *C4.5* tanpa fitur seleksi memiliki akurasi yaitu 86,8%. Hasil penelitian ini menunjukkan bahwa metode *C4.5* dengan penerapan fitur seleksi lebih baik dari *C4.5* tanpa fitur seleksi sehingga dapat diterapkan pada pemilihan penjual pakan ikan air tawar di Kabupaten Kendal.

Kata kunci: *Decision Tree, Klasifikasi, Pakan Ikan, Seleksi Fitur.*

1. PENDAHULUAN

Pakan ikan adalah salah satu variabel yang krusial dalam proses budidaya ikan, dimana dapat mengoptimalkan perkembangan ikan. Pertumbuhan

dan kelangsungan hidup ikan yang dibudidaya sangat dipengaruhi oleh ketersediaan pakan. Menurut Kementerian Kelautan dan Perikanan pada siaran pers tanggal 29 Agustus 2022, kebutuhan pakan ikan

di Indonesia pada tahun 2024 sebesar 22,65 juta ton yang dihitung dari target produksi perikanan budidaya [1]. Maka dari itu, pemerintah mengusahakan ketersediaan pakan ikan minimal 12-13 juta ton hingga tahun 2024. Dengan kebutuhan akan pakan yang begitu besar, membuktikan bahwa pembudidaya harus mulai memiliki penjual pakan ikan yang terpercaya agar dapat selalu memenuhi kebutuhan pakan untuk ikan yang dibudidayakannya.

Kabupaten Kendal adalah salah satu dari 29 kabupaten dan enam kota di Jawa Tengah [2]. Bentuk topografi Kabupaten Kendal terdiri dari wilayah pantai (wilayah utara), perbukitan, dan pegunungan (wilayah selatan) yang menyebabkan rata-rata suhu berkisar antara 25-29 °C. Dengan berbagai bentuk wilayah, Kabupaten Kendal tetap berupaya meningkatkan sektor perikanan melalui perikanan tangkap maupun budidaya. Menurut data Badan Pusat Statistik tentang budidaya ikan air tawar, pada tahun 2020 tercatat 228.608 pembudidaya [3] tersebar seluruh Kabupaten Kendal dengan berbagai jenis kolam mulai dari kolam permanen, kolam tanah, dan kolam terpal. Untuk ikan air tawar sendiri, mayoritas pembudidaya di Kabupaten Kendal membudidayakan ikan lele, nila, dan bawal dengan masing-masing volume produksi pada satuan kg yaitu 3.058.000, 74.800, dan 18.500. Jumlah tersebut merupakan angka yang besar dan perlunya kerjasama antara berbagai pihak untuk menyediakan berbagai fasilitas mulai dari pelatihan, pengairan, hingga ketersediaan obat dan pakan.

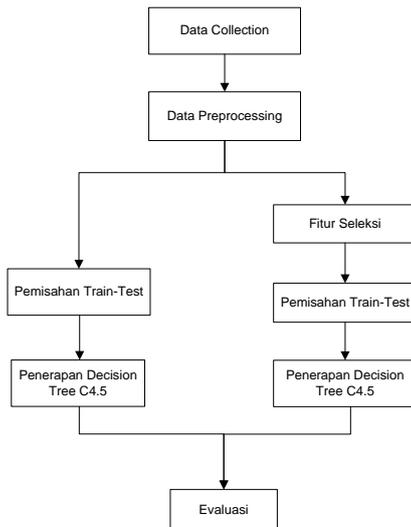
Secara umum, pakan ikan dibagi menjadi dua yaitu pakan alami dan pakan buatan. Pakan alami biasanya seperti plankton yang telah tersedia di kolam, sedangkan pakan buatan masih dibagi menjadi dua yaitu pakan alternatif dan pelet. Pakan alternatif merupakan pakan yang terbuat dari bahan pokok pakan utama seperti ampas tahu, maggot, dedaunan, dan lain-lain. Begitu pula dengan pelet, yaitu pakan komersil dengan komposisi yang telah disesuaikan dengan kebutuhan berdasarkan jenis dan umur ikan sehingga mampu memenuhi nutrisi bagi pertumbuhan dan kelangsungan hidup ikan. Untuk ikan lele, nila, dan bawal mayoritas pembudidaya biasanya mengkombinasikan pakan alternatif dan pelet untuk konsumsi ikan yang dibudidaya karena kedua pakan tersebut cukup terjangkau dan mudah ditemukan. Namun, dengan kombinasi tersebut ikan tidak cukup mendapat protein dan nutrisi secara optimal, hal ini membuat dilema pembudidaya karena untuk mendapatkan kualitas dan keuntungan yang maksimal membutuhkan konsistensi pakan dengan protein tinggi yang lebih optimal yaitu ditemukan di pelet. Permasalahan tersebut cukup membuat resah pembudidaya karena pembudidaya harus membeli pelet yang memakan 60-80% [4] biaya produksi dan pembudidaya juga harus menemukan penjual pakan ikan yang tepat sehingga pembudidaya dapat menghemat baik waktu, tenaga, serta biaya yang

dikeluarkan agar dapat lebih berkonsentrasi dan fokus untuk membudidayakan ikan.

Machine learning atau pembelajaran mesin adalah suatu metode yang mampu digunakan untuk memprediksi hasil di masa depan. Pembelajaran mesin berpusat pada peningkatan dan pembelajaran suatu sistem agar dapat menganalisis dari data yang sudah ada sebelumnya dimana merupakan bagian dari kecerdasan buatan [5]. Ciri pembeda dari pembelajaran mesin adalah terdapat pelatihan atau *training set*. Oleh karena itu, pembelajaran mesin memerlukan ketersediaan data yang dilatih (*training set*) dan diuji (*test set*) [6]. Data latih diperlukan untuk melatih algoritma *Decision Tree C4.5*, sedangkan *testing* data diperlukan untuk mengetahui kinerja *Decision Tree C4.5* yang sebelumnya telah dilakukan terlebih dahulu untuk mendapatkan data yang sebelumnya tidak diketahui. Sebagai bagian dari pembelajaran mesin, metode *C4.5* yaitu sebuah algoritma statistik yang dimanfaatkan untuk membangun pohon keputusan. Pohon keputusan atau *Decision Tree* adalah algoritma prediksi dan klasifikasi yang sangat kredibel dan populer. Algoritma *Decision Tree* mengkonversi sejumlah besar data menjadi aturan-aturan yang divisualisasikan dalam pohon keputusan. Aturan dalam bahasa alami akan lebih mudah dimengerti [7]. Pada penelitian ini, penggunaan *Decision Tree C4.5* dimanfaatkan sebagai metode dalam pemilihan penjual pakan ikan terbaik di Kabupaten Kendal. Penelitian ini berfokus pada klasifikasi penyederhanaan kriteria untuk penentuan pemilihan penjual pakan ikan dengan menggunakan fitur seleksi berbasis filter dengan variabel jumlah jenis pakan, harga, pemesanan, pengiriman, dan ketersediaan diskon, dimana untuk membantu pembudidaya ikan air tawar dalam menentukan penjual pakan ikan terbaik dari berbagai kriteria yang telah disebutkan sebelumnya. Diharapkan nilai *accuracy*, *recall*, dan *precision* pada analisis *Decision Tree C4.5* dengan atau tanpa penggunaan metode fitur seleksi mendapatkan hasil yang baik sehingga dapat menjadi acuan penelitian selanjutnya.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan data mining dan pemilihan fitur untuk menemukan pola atau kemungkinan dalam data. Singkatnya, ini adalah model penelitian yang akan digunakan untuk menentukan kriteria penjualan pakan ikan air tawar. Pemilihan fitur terbaik dari data yang ada menggunakan *SelectKBest* dari perpustakaan *scikit-learn*; melatih model *Decision Tree C4.5* dengan validasi silang 10 kali lipat; dan analisis confusion matrix untuk memverifikasi akurasi kinerja model. Gambar 1 menggambarkan enam tahapan penelitian ini: pengumpulan data, *preprocessing* data, pemilihan fitur, pembagian dataset menjadi pelatihan dan pengujian, implementasi algoritma, dan evaluasi.



Gambar 1. Alur tahapan penelitian

Sumber data yang digunakan dalam penelitian ini didapat dari *e-marketplace* dan wawancara langsung pada pembudidaya maupun penjual pakan ikan di Kabupaten Kendal.

2.1. Data Collection

Data yang dikumpulkan berjumlah 86 yang selanjutnya dipecah menjadi data *testing* dan data *training* dengan perbandingan 5:5. Variabel yang digunakan pada analisis pemrosesan yaitu jumlah jenis pakan, harga, pemesanan, pengiriman, pembayaran, ketersediaan diskon, dan jumlah penilaian. Penentuan kriteria variabel yang digunakan dirujuk dari penelitian yang dilakukan oleh Dickson [8], dimana dari 22 kriteria yang dipaparkan, penelitian ini hanya mengadopsi 7 variabel sebagai acuan. Selain itu, kriteria pemilihan variabel disurvei berdasarkan wawancara langsung terhadap pembudidaya agar sesuai dengan kebutuhan, meskipun terdapat beberapa faktor lain yang juga mendukung pengambilan keputusan pembudidaya pada pemilihan penjual pakan ikan. Berikut adalah deskripsi variabel dan dataset (Tabel 1) yang digunakan dalam penelitian ini.

1. Jenis_Pakan = jumlah jenis pakan yang dijual pada satu took
2. Harga = harga dari setiap jenis pakan yang dijual
3. Pemesanan = pilihan metode pemesanan yang ditawarkan
4. Pengiriman = pilihan metode pengiriman yang ditawarkan
5. Pembayaran = pilihan metode pembayaran yang ditawarkan
6. Diskon = ketersediaan diskon yang mungkin didapatkan
7. Jumlah rating = jumlah rating toko yang diambil dari *e-marketplace*
8. Rekomendasi = direkomendasikan atau tidak direkomendasikan.

Tabel 1. Dataset

Fitur	Nilai	Nilai	Nilai	Nilai
-------	-------	-------	-------	-------

Jumlah jenis pakan	3	1	1
Harga	Standar pasar	Standar pasar	Dibawah standar pasar
Pemesanan	Langsung ke toko	Langsung ke toko	Online
Pengiriman	Diantar dan diambil sendiri	Diambil sendiri	Diantar dan diambil sendiri
Pembayaran	Tunai Ongkir dan potongan harga	Tunai	Tunai
Diskon	Diskon	Diskon	Ongkir
Jumlah penilaian	3600	1221	2500
Rekomendasi	1	0	0

2.2. Data Preprocessing

Data preprocessing merupakan bagian dari tahapan yang paling penting dari pengolahan data untuk mempersiapkan data menjadi bentuk yang sesuai dengan prosedur dan tetap menjaga performa analisis data yang dilakukan [9]. *Data preprocessing* bertujuan untuk mengurangi ukuran data, menemukan hubungan antar data, menormalkan data, menghilangkan *outlier*, dan mengekstrak fitur data. Peneliti melakukan *preprocessing* untuk menghasilkan data yang siap diolah dengan cara membersihkan data yang hilang, yaitu dengan cara menghapus nilai yang menyebabkan rancunya informasi [10]. Karena dataset penelitain terdapat nilai yang hilang sebesar 14%, maka perlakuan yang harus diberikan yaitu mencari rata-rata nilai sesuai atribut yang berelasi sehingga semua data akan terisi dengan nilai yang valid. Tahap *preprocessing* selanjutnya, untuk menghindari data sampel yang bersifat abnormal dari sebuah populasi data, maka peneliti menggunakan *matplotlib* untuk mengetahui seberapa banyak data abnormal. Selain itu, untuk mengetahui apakah ada data yang duplikat, data diidentifikasi untuk mendeteksi adanya data lebih dari satu sehingga perlu dihilangkan dan akan menyisakan hanya satu data yang tersimpan.

2.3. Pemisahan Train-Test

Pemisahan *Train-Test* yaitu pembagian dataset menjadi *testing set* dan *training set*, atau dengan kata lain data yang dimanfaatkan pada tahapan *training* dan *testing* merupakan data set yang tidak sama. Metode pemisahan *train-test* akan menghasilkan nilai prediksi yang lebih tepat untuk data baru atau data yang belum dilatih [11,12]. Karena pengujian data tidak dimanfaatkan dalam pelatihan model, dimana model tidak peduli hasil data yang didapat [12]. Ini disebut pengujian di luar sampel. Data dalam penelitian ini akan dibagi 50% untuk data *training* dan 50% untuk data *testing*. Pembagian data ini bertujuan untuk memisahkan data menjadi data *test* dan data *train* agar dapat dievaluasi lebih spesifik.

2.4. Fitur Seleksi

Salah satu metode praklasifikasi yang biasa digunakan dalam pembelajaran mesin untuk meningkatkan kinerja pelatihan dan mereduksi data adalah pemilihan fitur. Sebelum proses pembangunan model klasifikasi, seleksi fitur dimanfaatkan sebagai pemilihan subset dari fitur yang relevan dan membuang fitur yang berlebihan, *redundant*, dan tidak diinginkan [13]. Pada penelitian ini, tahap pemilihan fitur memanfaatkan metode filter, dimana filter berfokus pada karakteristik umum dan statistik dari dataset dan tidak bergantung pada metode pembelajaran apapun [14]. Metode filter memiliki sifat murah secara komputasi dan kapasitas generalisasi yang baik karena independensinya dari algoritma yang digunakan.

2.5. Univariate Feature Selection

Kategori metode filter dalam proses pemilihan fitur meliputi pemilihan *fitur univariat* (*SelectKBest*). *SelectKBest* adalah algoritma pemilihan fitur untuk meningkatkan akurasi atau kinerja prediksi pada kumpulan data dimensi tinggi. *SelectKBest* bertujuan untuk memilih fitur input dengan membandingkan hasilnya ketika disertakan dalam fungsi tertentu [15]. *SelectKBest* adalah bagian dari tahapan seleksi fitur *univariat* yang memilih fitur terbaik berdasarkan uji statistik univariat atau uji ANOVA. Uji statistik dapat digunakan untuk mengidentifikasi fitur-fitur dengan hubungan paling kuat dengan variabel keluaran. *SelectKBest* menghapus semua kecuali fitur yang memiliki skor tertinggi. *SelectKBest* memilih fitur K teratas dengan relevansi terbesar dengan variabel target [16].

2.6. Pemodelan Data

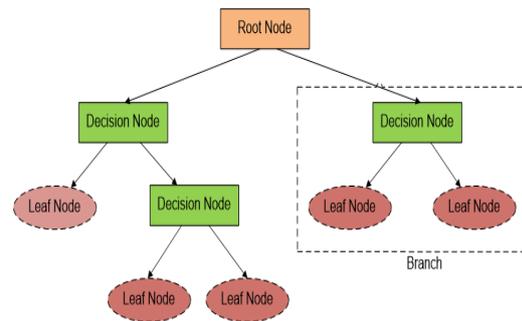
1) Klasifikasi

Klasifikasi [17] disebut algoritma supervisi yang menggunakan data pelatihan (pengamatan, pengukuran, dll.) dimana diatur dengan adanya atribut label. Klasifikasi berusaha untuk mengelompokkan kelas yang dituju dengan memaksimalkan perolehan presisi. Metode klasifikasi mencoba mendapatkan kemiripan antara atribut masukan dengan atribut keluaran guna menghasilkan model yang termasuk dalam tahapan pelatihan [18]. Klasifikasi membantu mengelompokkan atau mengkategorisasikan suatu data berdasarkan persamaan dan perbedaan yang telah dianalisis sebelumnya.

2) Decision Tree

Pohon keputusan digunakan untuk memproses data bertipe diskrit dan numerik, menangani atribut yang hilang, dan mendapatkan aturan yang mudah dipahami [19]. Pohon Keputusan dapat dimanfaatkan dalam mengeksplorasi data dan mengungkap hubungan tersembunyi antara sejumlah kandidat variabel input dan variabel target [20]. Penggunaan metode *decision tree* akan optimal jika data yang

diklasifikasikan mendapat hasil yang sesuai dan berisi jumlah minimum node [21]. Berikut adalah ilustrasi struktur dari *decision tree*.



Gambar 2. Struktur *decision tree*

3) Algoritma C4.5

C4.5 adalah algoritma pemrosesan segmentasi dan memprediksi data yang akan menjadi keputusan. Proses klasifikasi pada algoritma ini menghasilkan cabang sedangkan *leaf* merupakan kelas atau segmen [22]. Peneliti menggunakan metode C4.5 dikarenakan memiliki karakter data yang diklasifikasikan dapat dipahami dengan mudah mulai dari struktur *decision tree* maupun dalam struktur aturan *If-Then*, sehingga mudah untuk mengekstrak informasi mengenai data yang relevan. Algoritma *Decision Tree* mengkonversi sejumlah besar data menjadi aturan-aturan yang divisualisasikan dalam pohon keputusan dengan melakukan percabangan dari setiap nilai pada atribut *root*, membagi kasus menjadi cabang, dan mengulangi proses tersebut sampai semua kasus pada cabang memiliki nilai yang sama [19]. Berikut adalah cara kerja dari metode C4.5:

1. Mengumpulkan *data training* yang telah diklasifikasikan ke dalam kelas yang berbeda.
2. Pilih akar sebagai akar pertama dengan menghitung nilai *gain* dan menjadikan nilai *gain* tertinggi sebagai akar pertama. Namun, sebelum itu harus menghitung nilai *entropy*. Persamaan algoritma C4.5 secara matematis dapat ditulis sebagai berikut:

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

Keterangan:

S = Himpunan khusus

n = Jumlah partisi S

P_i = Proporsi S_i terhadap S

3. Selanjutnya, hitung nilai *gain* dengan persamaan sebagai berikut:

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

S = Himpunan khusus

A = Atribut

n = Jumlah partisi S

S_i = Jumlah partisi ke i

- Ulangi tahapan ke 2 dan 3 hingga seluruh *record* memiliki kelas yang sama, tidak ada *recod* yang dipartisi kembali, dan tidak ada cabang *record* yang kosong [19].

2.7. Confusion Matrix

Confusion matrix merupakan alat pengukur kinerja dimana dapat menghitung akurasi dalam konsep *data mining*. Rumus ini menghitung empat keluaran yaitu *accuracy*, *precision*, *recall*, dan tingkat kesalahan. Rasio kejadian yang dikenali dengan benar dengan jumlah total kejadian diketahui sebagai *accuracy*, proporsi kasus dengan hasil positif disebut *precision*, proporsi kasus positif yang teridentifikasi dengan benar disebut sebagai *recall*, tingkat kesalahan adalah jumlah dari semua kasus yang salah diidentifikasi [23]. Tabel 2 adalah model *confusion matrix*.

Tabel 2. Model *Confusion Matrix*

	+	-
+	<i>True positives</i>	<i>False negatives</i>
-	<i>False positives</i>	<i>True negatives</i>

3. HASIL DAN PEMBAHASAN

3.1. Pemodelan dengan Decision Tree C4.5

Penelitian ini bertujuan menganalisis dan menyederhanakan klasifikasi dalam penentuan pemilihan penjual pakan ikan bagi pembudidaya di daerah Kabupaten Kendal dengan menggunakan fitur seleksi berbasis filter dan nantinya kan dibandingkan dengan model tanpa penggunaan fitur seleksi agar mendapat algoritma dengan performa yang optimal sehingga pembudidaya mendapatkan manfaat yaitu dapat membeli pakan ikan dengan harga terjangkau dengan pelayanan yang unggul, hasilnya pembudidaya dapat memberikan pakan yang tinggi nutrisi pada ikan yang dibudidayakan dan diharapkan dapat mengoptimalkan hasil produksi yang didapat.

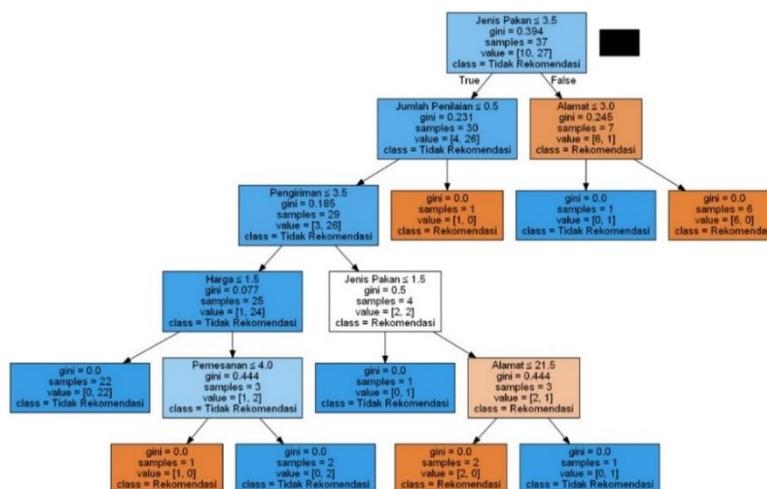
Pada penelitian ini, tahap pembersihan dan transformasi data menghasilkan dataset dengan 7

variabel atau fitur *input*. Untuk mengurangi variabel prediktor yang tidak perlu, pemilihan fitur digunakan untuk mengidentifikasi *input* atau variabel prediktor yang lebih signifikan, yang dapat meningkatkan akurasi model secara keseluruhan. Hanya tiga fitur yang diidentifikasi sebagai variabel prediktor yang dianggap signifikan, sedangkan yang lainnya tidak dipakai. Tebel 4 adalah urutan fitur yang direkomendasikan untuk dipakai dari skor tertinggi hingga terendah. Karena hanya diambil 3 fitur dengan nilai tertinggi, maka urutan fitur inilah yang memiliki peran penting dalam dataset, yaitu jenis pakan, jumlah penilaian, dan pengiriman. Tiga fitur yang sudah didapat nantinya akan dimodelkan menggunakan Algoritma *Decision Tree C4.5*.

Tabel 4. Seleksi Fitur

Fitur	Skor
Jenis Pakan	30.332362
Jumlah Penilaian	5.250000
Pengiriman	4.767086
Pemesanan	4.155844
Harga	1.749339
Diskon	0.329407
Alamat	0.284055

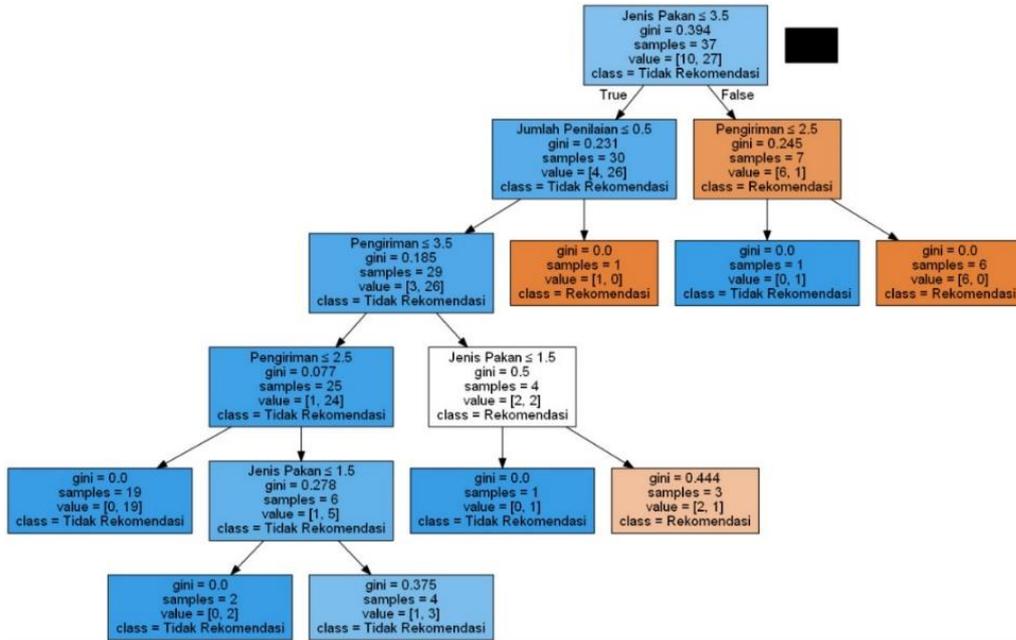
Dalam analisis pemrosesan data yang dilakukan dengan mengadopsi metode *data mining* untuk membuat klasifikasi dari dataset yang sebelumnya telah melalui *preprocessing* dan transformasi data, dimana keseluruhan data yang digunakan yaitu sebanyak 86 dimana dibagi menjadi data *training* yang berjumlah 43 dan data *testing* yang berjumlah 43. Setelah pemrosesan dilakukan maka tahap selanjutnya yaitu mengetahui nilai *accuracy*, *precision*, dan *recall* dengan melakukan pengukuran kinerja menggunakan *confusion matrix*, hal ini agar pemodelan menghasilkan data yang valid. Gambar 3 memvisualisasikan hasil pohon keputusan dari analisis dengan algoritma *C4.5* tanpa implementasi fitur seleksi berbasis filter yang telah dilakukan. Pohon keputusan yang dihasilkan cukup rimbun dan sudah mewakili aturan untuk penentuan kriteria penjual pakan ikan air tawar.



Gambar 3. Pohon Keputusan tanpa filter

Untuk menyederhanakan pohon keputusan maka akan divisualisasikan pohon keputusan dengan fitur seleksi berbasis filter. Gambar 4 merupakan visualisasi penyederhanaan pohon keputusan dengan menggunakan tiga fitur yaitu jenis pakan, jumlah

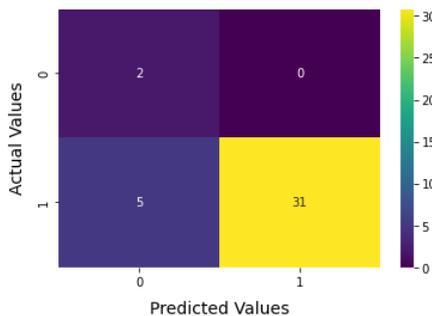
penilaian, dan pengiriman. Pohon keputusan yang dihasilkan tetap rimbun dan dapat mewakili klasifikasi dalam penentuan kriteria penjual pakan namun lebih mudah dipahami sehingga ketika diterapkan pengguna tidak merasa kebingungan.



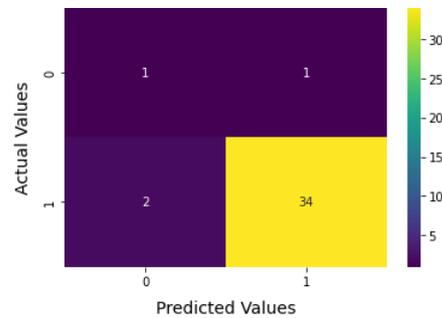
Gambar 4. Pohon Keputusan Dengan Filter

3.2. Evaluasi Model

Tahap selanjutnya yaitu untuk mengukur kinerja dari hasil klasifikasi yang telah dilakukan. Peneliti menggunakan *confusion matrix* untuk menunjukkan dan menguji tingkat kebenaran bagaimana sebuah model dapat menghasilkan klasifikasi. Gambar 5 dan 6 menunjukkan hasil kinerja dari kedua model yang dihasilkan dalam evaluasi klasifikasi. Gambar 5 menggunakan algoritma *C4.5* tanpa adanya fitur seleksi dan Gambar 6 menggunakan algoritma *C4.5* dengan penambahan fitur seleksi berbasis filter. Kedua model telah diimplementasikan dengan validasi silang 10 kali lipat. *Cross-validation* merupakan salah satu metode teknik *data mining* yang bertujuan untuk mendapatkan hasil akurasi yang maksimal.



Gambar 5. Confusion Matrix tanpa fitur seleksi



Gambar 6. Confusion Matrix dengan fitur seleksi

Setelah mendapatkan visualisasi pengukuran *confusion matrix*, untuk mendapatkan hasil akurasi, presisi, dan *recall* maka bisa menggunakan persamaan pada Tabel 2. Tujuan dilakukannya analisis nilai *confusion matrix* adalah agar dapat diketahui secara spesifik kinerja kedua model tersebut. Berikut adalah hasil perhitungan *confusion matrix* untuk algoritma *C4.5* dengan dan tanpa penambahan fitur seleksi.

Table 3. Evaluasi model

	Akurasi	Presisi	Recall
C4.5 tanpa fitur seleksi	86,8%	100%	86%
C4.5 dengan fitur seleksi	92,1%	97%	94%

Dapat dilihat pada Tabel 3 bahwa secara keseluruhan kinerja algoritma *C4.5* dengan penerapan fitur seleksi lebih stabil daripada kinerja algoritma *C4.5* tanpa penerapan fitur seleksi. Hal ini menunjukkan bahwa fitur seleksi lebih optimal

digunakan untuk merekomendasikan penjual pakan ikan kepada pembudidaya.

Karena telah terbukti bahwa algoritma *C4.5* dengan penerapan fitur seleksi lebih optimal digunakan dalam klasifikasi, maka untuk memastikan keakuratan nilai prediksi rekomendasi penjual pakan ikan atau tidak, peneliti mengujinya untuk mengetahui nilai prediksi positif dan negatifnya. Pengujian dilakukan untuk memprediksi apakah yang dikerjakan dapat berjalan sesuai fungsinya. Berikut perhitungan untuk uji prediksi:

1. Nilai prediktif positif:

$$PPV = \frac{TP}{(TP+FP)} = \frac{34}{(34+1)} = 0,97 \quad (3)$$

Jika hasil prediksi merupakan rekomendasi, maka nilai prediktif positif menunjukkan kemungkinan penjual pakan ikan tersebut memiliki karakteristik yang ditunjukkan oleh fitur tersebut. NPP sebesar 0,97 menunjukkan bahwa jika hasil prediksi menghasilkan rekomendasi, maka kemungkinan penjual benih ikan merekomendasikan sebesar 97%.

2. Nilai tes prediksi negatif :

$$NPV = \frac{TN}{(TN+FN)} = \frac{1}{(1+2)} = 0,333 \quad (4)$$

Jika hasil prediksi tidak menjadi rekomendasi, maka nilai prediksi negatif menunjukkan kemungkinan penjual pakan ikan memiliki kesamaan karakteristik. Jika hasil prediksi tidak direkomendasikan maka peluang penjual pakan ikan tidak direkomendasikan sebesar 34%. Secara keseluruhan algoritma *C4.5* dengan penambahan fitur seleksi berbasis filter lebih baik dalam menganalisis dan menyederhanakan kriteria penjual pakan ikan yang diharapkan dapat memberikan prediksi apakah penjual pakan ikan masuk dalam kategori rekomendasi atau tidak sesuai dengan kebutuhan pembudidaya ikan air tawar di Kabupaten Kendal. Peneliti berharap dengan adanya prediksi ini pembudidaya dapat menghindari penjual pakan yang tidak dianjurkan yang dapat merugikan pembudidaya itu sendiri.

4. DISKUSI

Pada sebuah penelitian yang dilakukan oleh [24], bertujuan menganalisis dan meninjau bagaimana metode filter yang berbeda berfungsi, untuk membandingkan kinerjanya sehubungan dengan *runtime* dan akurasi prediktif, dan untuk memberikan panduan untuk aplikasi. Berdasarkan 16 set data klasifikasi dimensi tinggi, 22 metode filter dianalisis sehubungan dengan *runtime* dan akurasi saat digabungkan dengan metode klasifikasi. Kesimpulannya metode filter tidak mengungguli penggunaan metode klasifikasi tanpa filter, namun metode filter direkomendasikan karena memiliki kinerja baik pada sekumpulan data yang diolah.

Selain itu, pada penelitian lain [25] yang menjelaskan beberapa dasar konsep yang terkait dengan aplikasi medis dan memberikan beberapa informasi yang diperlukan dalam pemilihan fitur. Peneliti meninjau metode pemilihan fitur yang dikembangkan dan diterapkan dalam masalah medis, meliputi bidang penelitian produktif seperti pencitraan medis, pemrosesan sinyal biomedis, dan *microarray* DNA analisis data. Sebuah studi kasus dari dua aplikasi medis yang menyertakan data pasien aktual digunakan untuk menunjukkan kesesuaian penerapan metode pemilihan fitur dalam masalah medis dan untuk menggambarkan bagaimana metode tersebut bekerja dalam skenario dunia nyata.

Dari kedua penelitian diatas menunjukkan bahwa kedua metode fitur seleksi berbasis filter memiliki hasil klasifikasi yang cukup baik. Namun, belum ditemukan penelitian pada bidang perikanan sekaligus dengan mengimplementasikan metode *Decision Tree C4.5*. Maka dari itu, tujuan yang akan dicapai pada penelitian ini yaitu untuk menganalisis dan menyederhanakan kriteria pemilihan penjual pakan ikan memanfaatkan metode *Decision Tree C4.5* dengan fitur seleksi berbasis filter dan kemudian hasilnya akan dibandingkan dengan metode tanpa fitur seleksi.

5. KESIMPULAN

Hasil analisis pencarian model terbaik antara *C4.5* dengan dan tanpa penerapan fitur seleksi berbasis filter dari data penjual pakan ikan air tawar di Kabupaten Kendal diperoleh hasil bahwa algoritma *C4.5* dengan penerapan fitur seleksi berbasis filter terbukti lebih baik dan optimal untuk menentukan kriteria penjual pakan ikan air tawar dibandingkan dengan menggunakan algoritma *C4.5* tanpa penerapan fitur seleksi, hal ini dikarenakan nilai performansi yang diperoleh *C4.5* dengan penerapan fitur seleksi lebih stabil dengan nilai akurasi 92,1%, presisi 97%, dan *recall* 94%. Kriteria yang digunakan dalam menentukan pemilihan penjual ikan air tawar yaitu jumlah jenis pakan, jumlah penilaian, dan pengiriman sehingga dapat diterapkan oleh pembudidaya untuk mengoptimalkan proses penentuan penjual pakan ikan yang akan dipilih. Selanjutnya penelitian ini perlu dilakukan perbaikan terhadap penggunaan data yang masih terlalu sedikit sehingga memerlukan lebih banyak data yang mewakili penjual pakan se-Kabupaten Kendal sehingga dapat menjadi acuan bagi pembudidaya dalam memperoleh penjual pakan ikan air tawar yang tepat.

DAFTAR PUSTAKA

- [1] Badan Riset Dan Sumber Daya Manusia Kelautan Dan Perikanan, "Kkp Gencarkan Pelatihan Guna Wujudkan Ketersediaan Pakan Ikan Mandiri," 2022. <https://kkp.go.id/brsdm/artikel/44444-kkp->

- gencarkan-pelatihan-guna-wujudkan-ketersediaan-pakan-ikan-mandiri (accessed Oct. 2, 2022).
- [2] Badan Pusat Statistik, "Jumlah Kecamatan Menurut Kabupaten/Kota di Provinsi Jawa Tengah 2019-2021," 2022. <https://jateng.bps.go.id/indicator/101/457/1/jumlah-kecamatan-menurut-kabupaten-kota-di-provinsi-jawa-tengah.html> (accessed Oct. 2, 2022).
- [3] Statistik KKP, "Produksi Perikanan Per Kabupaten Kota," 2022. https://statistik.kkp.go.id/home.php?m=prod_kabkota&i=2#panel-footer, (accessed Oct. 2, 2022).
- [4] Yunaidi, P. A. Rahmanta, A. Wibowo. "Aplikasi pakan pelet buatan untuk peningkatan produktivitas budidaya ikan air tawar di desa Jerukagung Srumbung Magelang," *Jurnal Pemberdayaan: Publikasi Hasil Pengabdian kepada Masyarakat*, vol. 3, no. 1, pp. 45-54, 2019, doi : <https://doi.org/10.12928/jp.v3i1.621>
- [5] A. Prasetyo, Salahuddin, Amirullah, "Prediksi Produksi Kelapa Sawit Menggunakan Metode Regresi Linier Berganda," *Jurnal Infomedia: Teknik Informatika, Multimedia & Jaringan*, vol. 6, no. 2, pp. 76-80, 2021, doi : <http://dx.doi.org/10.30811/jim.v6i2.2343>.
- [6] Wijaya, H. Derajad, S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *Jurnal Informatika*, vol. 7, no. 1, pp. 1-7, 2020, doi : <https://doi.org/10.31294/ji.v7i1.6203>.
- [7] T. H. Zebua, F. Riandari, "Implementation Of Data Mining With C4. 5 Algorithm For Determining The Home Industry Product Marketing Strategy," *Journal of Intelligent Decision Support System (IDSS)*, vol. 4, no. 4, pp. 113-121, 2021, doi : <https://doi.org/10.35335/idss.v4i4.37>.
- [8] G. W. Dickson, "An analysis of vendor selection systems and decisions," *Journal of purchasing*, vol. 2, no.1, pp. 5-17, 1966, doi : <https://doi.org/10.1111/j.1745-493X.1966.tb00818.x>.
- [9] I. M. A. Agastya, "Pengaruh stemmer bahasa Indonesia terhadap peforma analisis sentimen terjemahan ulasan film," *Jurnal Tekno Kompak*, vol. 12, no. 1, pp. 18-23, 2018, doi : <https://doi.org/10.33365/jtk.v12i1.70>.
- [10] P. S. Raja, K. Sasirekha, and K. Thangavel, "A novel fuzzy rough clustering parameter-based missing value imputation," *Neural Computing and Applications*, vol. 32, pp. 10033-10050, 2020, doi : <https://doi.org/10.1007/s00521-019-04535-9>.
- [11] C. Thapa, P. C. M. Arachchige, S. Camtepe, "Splitfed: When federated learning meets split learning," *In Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, vol. 36, no. 8, pp. 8485-8493, doi : <https://doi.org/10.1609/aaai.v36i8.20825>.
- [12] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, "Introduction to machine learning, neural networks, and deep learning," *Translational Vision Science & Technology*, vol. 9, no. 2, pp. 14-14, 2020, doi : <https://doi.org/10.1167/tvst.9.2.14>.
- [13] D. Ariyoga, "Perbandingan Metode Seleksi Fitur Filter, Wrapper, Dan Embedded Pada Klasifikasi Data Nirs Mangga Menggunakan Random Forest Dan Support Vector Machine (Svm)," Universitas Islam Indonesia, 2022.
- [14] S. Lu, S. Shen, J. Huang, M. Dong, J. Lu, W. Li, "Feature selection of laserinduced breakdown spectroscopy data for steel aging estimation," *Spectrochimica Acta - Part B Atomic Spectroscopy*, vol. 150, pp. 49–58, 2018, doi : <https://doi.org/10.1016/j.sab.2018.10.006>.
- [15] T. Desyani, A. Saifudin, Y. Yulianti. "Feature Selection Based on Naive Bayes for Caesarean Section Prediction," *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 879. no. 1, doi : [10.1088/1757-899X/879/1/012091](https://doi.org/10.1088/1757-899X/879/1/012091).
- [16] Pritalia, G. L. "Analisis Komparatif Algoritme Machine Learning pada Klasifikasi Kualitas Air Layak Minum," *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol. 2, no. 1, pp. 43-55, 2022, doi : <https://doi.org/10.24002/konstelasi.v2i1.5630>.
- [17] P. Ardiyansyah, Rahayuningsih, R. Maulana, "Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner," *Jurnal Khatulistiwa Informatika*, vol. 6, no. 1, 2018, doi : <https://doi.org/10.31294/jki.v6i1.3799>.
- [18] S. Faisal, "Klasifikasi data minning menggunakan algoritma c4. 5 terhadap kepuasan pelanggan sewa kamera cikarang," *Techno Xplore: Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 4, no.1, pp. 38-45, 2019, doi : <https://doi.org/10.36805/technoxplore.v4i1.541>.
- [19] Y. I. Kurniawan, A. Fatikasari, M. L. Hidayat, M. Waluyo, "Prediction For Cooperative Credit Eligibility Using Data

- Mining Classification With C4. 5 Algorithm," *Jurnal Teknik Informatika (JUTIF)*, vol. 2, no. 2, pp. 67-74, 2021, doi: <https://doi.org/10.20884/1.jutif.2021.2.2.49>
- [20] M. Ula, A. F. Ulva, M. Mauliza, M. A. Ali, Y. R. Said, "Application Of Machine Learning In Determining The Classification Of Children's Nutrition With Decision Tree," *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 5, pp. 1457-1465, 2022, doi : <https://doi.org/10.20884/1.jutif.2022.3.5.599>
- [21] B. Kaminski, M. Jakubczyk, P. Szufel. "A framework for sensitivity analysis of decision trees," *Central European Journal of Operations Research*, vol. 26, pp. 135–159, 2018, doi : <https://doi.org/10.1007/s10100-017-0479-6>.
- [22] J. Eska, "Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 2, no. 2, pp. 9-13, 2018, doi : [10.31227/osf.io/x6svc](https://doi.org/10.31227/osf.io/x6svc).
- [23] O. Ahmed, A. Brifcani, "Gene Expression Classification Based on Deep Learning," *4th Scientific International Conference Najaf (SICN)*, 2019, vol. 2019, pp. 145-149, doi : [10.1109/SICN47020.2019.9019357](https://doi.org/10.1109/SICN47020.2019.9019357) .
- [24] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, M. Lang, "Benchmark for filter methods for feature selection in high-dimensional classification data," *Computational Statistics & Data Analysis*, vol. 143, pp. 106839, 2020, doi : <https://doi.org/10.1016/j.csda.2019.106839>.
- [25] B. Remeseiro, V. B. Canedo, "A review of feature selection methods in medical applications," *Computers in biology and medicine*, vol. 112, pp. 103375, 2019, doi : <https://doi.org/10.1016/j.compbiomed.2019.103375>..