Analisis Data mining dengan Metode C.45 pada Klasifikasi Kenaikan Rata-Rata Volume Perikanan Tangkap

Muhammad Ridho Matondang¹, Muhammad Ridwan Lubis², Heru Satria Tambunan³

^{1,3}STIKOM Tunas Bangsa Pematangsiantar, Sumatera Utara, Indonesia

²AMIK Tunas Bangsa Pematangsiantar, Sumatera Utara, Indonesia

Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara

¹mhdridho738@gmail.com, ²ridwanlubis@amiktunasbangsa.ac.id,

³herusatria@stikomtunasbangsa.ac.id

Abstract

Increasing the amount of demand for natural resource needs is increasing. One of them is natural resources in the sea and coast. The current condition of capture fisheries in Indonesia is not yet optimal. This is indicated by the increase in the volume of capture fisheries production which is very slow. The purpose of this study is to make data classification for the prediction of the average volume increase in capture fisheries with data mining techniques. Data mining techniques are applied to determine the data patterns of the capture fisheries dataset, so the results of the classification can be applied to evaluate the factors that affect the volume of capture fisheries. The classification algorithm used is C45. The results of the classification were tested with rapidminer in classifying data. The level of performance is indicated by the accuracy value. The accuracy value is obtained by testing the results of the classification of training data and testing data. Comparison of accuracy values between the algorithms used can be seen the best algorithm in making the classification of capture fisheries data.

Keywords : Capture fisheries, Decision Tree, C45

Abstrak

Peningkatan jumlah permintaan terhadap kebutuhan sumber daya alam semakin meningkat. Salah satunya sumber daya alam yang berada di laut dan pesisir. Kondisi perikanan tangkap di Indonesia saat ini belum optimal. Hal tersebut ditunjukkan dengan peningkatan volume produksi perikanan tangkap yang sangat lambat. Tujuan penelitian ini membuat klasifikasi data untuk prediksi kenaikan volume rata-rata perikanan tangkap dengan teknik data mining. Teknik data mining diterapkan untuk mengetahui pola data dari dataset perikanan tangkap, sehingga hasil klasifikasi dapat diterapkan untuk mengevaluasi faktor-faktor yang berpengaruh terhadap volume perikanan tangkap. Algoritma klasifikasi yang digunakan yaitu C45. Hasil klasifikasi diuji dengan rapidminer dalam mengklasifikasikan data. Tingkat performance ditunjukkan dengan nilai akurasi. Nilai akurasi tersebut diperoleh dengan pengujian hasil klasifikasi terhadap data training dan data testing. Perbandingan nilai akurasi antar algoritma yang digunakan dapat diketahui algoritma terbaik dalam membuat klasifikasi data perikanan tangkap.

Kata Kunci : Perikanan tangkap, Decision Tree, C45

1. Pendahuluan

Indonesia memiliki potensi yang besar dalam budidaya perikanan yang belum dimanfaatkan dan dikelola secara optimal. Potensi ekonomi tersebut didukung pula oleh tingkat keanekaragaman hayati ikan yang tinggi dan sangat signifikan untuk memenuhi kebutuhan ikan didalam negeri dan memasok permintaan ikan dipasar dunia sehingga dapat menjadi sumber devisa bagi negara. Dengan satu paket

teknologi tertentu, ternyata perairan umum dapat dijadikan sebagai media budidaya ikan secara intensif. Teknologi tersebut adalah dengan menggunakan kantong jaring terapung, keramba atau sistem pagar. Ternyata dengan cara itu diperoleh produksi yang tinggi dan memberikan keuntungan yang tinggi dan tentunya memberikan keuntungan yang tinggi bagi petani ikan. Ikan merupakan salah satu sumber zat gizi penting bagi proses.

Salah satu cara dalam pemanfaatan sumber daya alam di laut dan pesisir yaitu dengan perikanan tangkap. Kondisi perikanan tangkap di Kabupaten Simalungun saat ini belum optimal, beberapa daerah mengalami kemunduran produktivitas hasil perikanan tangkap yang berdampak pada minimnya kontribusi terhadap peningkatan produktivitas perikanan tangkap. Sedangkan volume produksi perikanan secara keseluruhan mengalami sedikit peningkatan dari tahun ke tahun. Evaluasi kenaikan volume produksi perikanan dapat dilakukan dengan mengevaluasi faktor-faktor yang mempengaruhinya. Faktor yang sering dialami sebagian besar petani ikan ialah keterbatasan modal untuk menyediakan sarana produksi berupa bibit ikan dan pakan/pellet. penelitian ini bertujuan membuat klasifikasi data untuk mengetahui faktor yang berpengaruh terhadap kenaikan volume perikanan tangkap dengan menerapkan teknik data mining.

Usaha yang dilakukan untuk mencegah stagnasi tersebut adalah dengan cara menganalisa faktor terjadinya penurunan volume ikan tangkapan yaitu dengan mengklasifikasi hasil rata rata ikan tangkapan dari beberapa daerah di Kabupaten Simalungun. untuk mengelompokkan kenaikan volume perikanan tangkap perlu menerapkan teknik data mining. Menurut Rofika dalam jurnal JIKO *Data mining* adalah suatu kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menentukan keteraturan, pola atau hubungan dalam set data berukuran besar. Salah satu tugas utama dalam *data mining* adalah pengklasifikasian dimana data yang diklasifikasi belum mempunyai contoh kelompok. "Data Mining merupakan proses ekstraksi data menjadi informasi yang sebelumnya belum tersampaikan, dengan teknik yang tepat proses data mining akan memberikan hasil yang optimal"[1]. Algoritma yang digunakan dalam pengelompokan tersebut menggunakan algoritma C45. Algoritma C4.5 adalah algoritma klasifikasi data bertipe pohon keputusan [2]. Pohon keputusan Algoritma C4.5 dibangun dengan tahapan yang meliputi pemilihan atribut sebagai node awal, membentuk cabang untuk setiap nilai dan mempartisi kasus dalam cabang. Tahapan-tahapan ini akan diulangi untuk setiap cabang sampai semua kasus pada cabang habis terpartisi. Dari penyelesaian tersebut maka akan didapatkan beberapa rule. Diharapkan penelitian ini dapat membantu pihak petani ikan dalam meningkatkan hasil ikan tangkapan mereka.

2. Metodologi Penelitian

2.1. Data Mining

Data Mining Merupakan suatu rangkaian proses yang dapat dibagi menjadi beberapa tahap yang mana Tahap-tahap tersebut bersifat *interaktif* di mana pemakai terlibat langsung atau dengan perantara *knowledge base* [3]. *Data Mining* sering juga disebut Knowledge discovery in database (KDD) merupakan meliputi kegiatan pengumpulan, pemakaian data historis menemukan pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan [4].

2.2. Algoritma C4.5

Algoritma C4.5 adalah algoritma yang digunakan untuk menghasilkan sebuah pohon keputusan yang dikembangkan oleh Ross quinlan .Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai *Gain* tertinggi [5] berdasarkan

nilai *Entropy* atribut tersebut sebagai poros atribut klasifikasi. Pada tahapannya algoritma C4.5 memiliki 2 prinsip kerja yaitu: Membuat pohon keputusan, dan membuat aturan-aturan (*rule model*). Aturan aturan yang terbentuk dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if then*. Beberapa tahapan yang ada untuk membuat pohon keputusan dengan algoritma C4.5 yaitu [6] :

- Menyiapkan data training. Data ini diambil dari data yang sudah pernah ada sebelumnya dan sudah dikelompokkan ke dalam kelas tertentu.
- Setelah itu tentukan *node* awal dari pohon. Pilih *node* awal dari atribut, dengan cara menghitung nilai *Gain* dari semua atribut, yang menjadi *node* awal adalah nilai *Gain* yang paling tinggi. Sebelum menentukan nilai *Gain*, terlebih dahulu hitung nilai *Entropy* total dengan menggunakan rumus :

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

Keterangan:

S= himpunan kasus

n= jumlah partisi s

Pi= proporsi Si terhadap S

- Setelah itu tentukan nilai *Gain* menggunakan rumus:

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

Keterangan :

S = Himpunan Kasus

A = Atribut

n = Jumlah Partisi Atribut A

| Si | = Jumlah Kasus pada partisi ke-i

| S | = Jumlah Kasus dalam S

- Setelah itu ulangi langkah ke-2 sampai semua record habis terpartisi dan mendapat kelas yang sama.
- Proses partisi akan berhenti jika :
 - Semua record dalam simpul N mendapat kelas yang sama.
 - Tidak ada atribut di dalam record yang dapat dipartisi lagi.
 - Tidak ada record di dalam cabang yang kosong.

Gambar 1. Model Pohon Keputusan

2.3. Rapidminer

Rapidminer adalah platform perangkat ilmu data yang dikembangkan oleh perusahaan bernama sama dengan yang menyediakan lingkungan terintegrasi untuk persiapan data pembelajaran mesin pembelajaran dalam pengembangan teks dan analisis prediktif. Memudahkan pengguna dalam menggunakan perangkat ini. Berbagai metode yang disediakan aplikasi Rapidminer antara lain klasifikasi, *Clustering* dan *Assosiasi* [8].

3. Hasil dan Pembahasan

Penelitian ini dilakukan menggunakan data direktorat ikan tangkapan Kabupaten Simalungun yang diambil dari situs resmi Badan Pusat Statistik atau <https://www.bps.go.id/>. Jenis data yang digunakan dalam penelitian ini merupakan data sekunder. Data sekunder adalah data yang diperoleh tidak dari sumbernya langsung melainkan sudah dikumpulkan oleh pihak lain dan sudah di olah serta memiliki keterkaitan dengan permasalahan yang diteliti.

Penelitian ini menggunakan Algoritma C45 dalam menentukan Faktor ikan tangkapan. Berdasarkan data yang digunakan maka tahap pertama yang harus dilakukan yaitu menghitung nilai *Entropy* dan *Gain*. Untuk menghitung nilai *Entropy* dan *Gain* diperlukan data awalnya dahulu.

Tabel 1. Tabel data awal

Kelompok	Jumlah Nelayan	Kondisi Cuaca	Lebar Jaring	Kondisi Air	Jumlah Trip	Output
1	10	cerah	60 m	Surut	4	menengah
2	10	Hujan lebat	50 m	Pasang	3	menengah
3	15	Hujan lebat	50 m	Pasang	2	menengah tinggi
4	15	cerah	40 m	Surut	2	menengah tinggi
5	15	cerah	40 m	Surut	2	menengah tinggi
6	15	cerah	50 m	Surut	3	menengah tinggi
7	15	Hujan lebat	60 m	Pasang	2	menengah tinggi
8	15	Hujan lebat	50 m	Pasang	2	menengah tinggi
9	20	cerah	50 m	Surut	3	menengah tinggi
10	20	Hujan lebat	60 m	Pasang	4	menengah tinggi
11	20	Hujan lebat	60 m	Pasang	4	menengah tinggi
12	20	Hujan lebat	60 m	Pasang	3	menengah tinggi
13	10	cerah	60 m	Surut	3	menengah
14	20	Hujan lebat	60 m	Pasang	3	menengah
15	20	Hujan lebat	60 m	Pasang	3	menengah
16	20	cerah	50 m	Surut	3	menengah tinggi
17	20	cerah	50 m	Surut	3	menengah tinggi
18	15	berangin	50 m	Berombak	3	menengah tinggi

Output atau target variabel memiliki dua kategori yaitu menengah Tinggi dan Menengah, maka digunakan teknik klasifikasi yang termasuk dalam pengelompokan data karena dalam klasifikasi terdapat target variabel kategori. Variabel yang akan menjadi keputusan adalah Menengah dan Menengah Tinggi. Karena Implementasi Algoritma C 4.5 dilakukan dengan cara menjadikan atribut Kenaikan rata-rata hasil ikan tangkapan dengan klasifikasi menengah dan menengah tinggi sebagai atribut tujuan. sedangkan atribut parameter adalah Jumlah nelayan, Kondisi cuaca, Lebar jaring, Kondisi air dan jumlah trip mencari ikan. Format data akhir pada tabel 4.1 didapat berdasarkan dari *attribute* yang sudah dikelompokkan atau diklasifikasi, misalkan data pada tabel 4.1 Jumlah Nelayan diklasifikasi kan menjadi 3 yaitu 10 orang, 15 orang dan 20 orang. Kondisi Cuaca diklasifikasikan menjadi Hujan lebat, cerah dan berangin. Lebar jaring diklasifikasikan menjadi 40m, 50m dan 60m. Kondisi Air diklasifikasikan menjadi Pasang, surut dan berombak. Jumlah trip diklasifikasikan menjadi 2 trip, 3 trip dan 4 trip.

Pohon keputusan dibuat setelah menghitung *Entropy* total, dengan *Entropy* masing- masing atribut dan menghitung *Gain* dan menentukan *Gain* tertinggi, hal ini dapat dilihat dari Tabel node 1 berikut ini :

Tabel 2. Tabel Perhitungan Node 1

Node 1	Faktor	Sub	Jumlah				<i>Entropy</i>	<i>Gain</i>
			Faktor	Menengah Tinggi	Menengah			
			18	13	5		0,852405	
	Jumlah nelayan	10	3	0	3		0	0,461056
		15	7	7	0		0	

	Kondisi Cuaca	20	8	6	2	0,880535	
		cerah	8	6	2	0,880535	-
		Hujan Lebat	9	6	3	0,959148	0,01852
		Berangin	1	1	0	0	
	Lebar Jaring	40 m	2	2	0	0	0,085311
		50 m	8	7	1	0,761551	
		60 m	8	4	4	0,964411	
	Kondisi Air	Pasang	9	6	3	0,959148	-
		Surut	8	6	2	0,880535	0,01852
		Berombak	1	1	0	0	
	Jumlah Trip	4	3	2	1	0,583876	0,193689
		3	10	6	4	1,010526	
		2	5	5	0	0	

Dari table 1 dapat diketahui bahwa pada baris total diketahui jumlah kasus (S) adalah 18, jumlah Menengah tinggi (S1) adalah 13, dan jumlah Menengah (S2) adalah 5. Perhitungan *Entropy* total pada table 1. dapat dihitung dengan menggunakan persamaan sebagai berikut:

$$Entropy (Total) = \left(-\frac{13}{18} * \log_2 \left(\frac{5}{18} \right) \right) + \left(-\frac{5}{18} * \log_2 \left(\frac{5}{18} \right) \right)$$

$$= 0,852405$$

Gambar 2. Node Akar 1

Dari gambar dapat dilihat bahwa atribut Jumlah nelayan memiliki 3 nilai yaitu jumlah nelayan 10 orang, Jumlah nelayan 15 orang, dan jumlah nelayan 20 orang. Dari tiga nilai tersebut masih memerlukan perhitungan lebih lanjut.

Sementara itu untuk perhitungan *Entropy* setiap atribut sama halnya dengan perhitungan mencari nilai *Entropy (total)*. Setelah nilai *Entropy* didapat selanjutnya mencari nilai *Gain* pada setiap atribut. Nilai *Gain* atribut *Gain* dihitung dengan menggunakan persamaan sebagai berikut :

$$Gain = 0.852405 - \left(\left(\frac{3}{18} * 0 \right) + \left(\frac{7}{18} * 0.880535 \right) + \left(\frac{6}{18} * 0 \right) \right)$$

$$Gain = 0,461056$$

Tabel 3. Tabel Perhitungan Node 1.2

Node 1.2	Faktor	Sub	Jumlah				
			Faktor	Menengah Tinggi	Menengah	Entropy	Gain
			8	6	2	0,880535	
	Jumlah nelayan, 20						
	Kondisi Cuaca	cerah	3	3	0	0	0,236385
		Hujan Lebat	5	3	2	1,030639	
		berangin	0	0	0	0	

	Lebar Jaring	40 m	0	0	0	0	0,236385
		50 m	3	3	0	0	
		60 m	5	3	2	1,030639	
	Kondisi Air	Pasang	5	3	2	1,030639	0,236385
		Surut	3	3	0	0	
		Berombak	0	0	0	0	
	Jumlah Trip	4	2	2	0	0	0,130535
		3	6	4	2	1	
		2	0	0	0	0

Gambar 3. Node Akar 1.2

Dari gambar dapat dilihat bahwa atribut Kondisi Cuaca memiliki 2 nilai yaitu Hujan Lebat dan Cerah. Dari nilai tersebut masih memerlukan perhitungan lebih lanjut.

Proses perhitungan dalam mencari nilai *Entropy* dan *Gain* sama dengan perhitungan untuk mencari nilai *Entropy* dan *Gain* pada node 1. Jumlah kasus yang digunakan disesuaikan dengan jumlah kasus setiap nilai yang ada pada atribut Jumlah Nelayan. Perhitungan akan terus berlanjut sampai semua atribut yang ada habis terpartisi.

Gambar 4. Pohon Keputusan Node 2.2.

Adapun aturan atau *rule* yang terbentuk berdasarkan pohon keputusan terakhir seperti pada gambar 4 di atas adalah sebagai berikut :

```

Jumlah Nelayan = 10 orang: menengah {menengah=3, menengah tinggi=0}
Jumlah Nelayan = 15 orang: menengah tinggi {menengah=0, menengah tinggi=7}
Jumlah Nelayan = 20 orang
|   Kondisi Cuaca = Hujan lebat
|   |   Jumlah Trip = 3 kali
|   |   |   Lebar Jaring = 60 m
|   |   |   |   Kondisi Air = Pasang : menengah {menengah=2, menengah tinggi=1}
|   |   Jumlah Trip = 4 kali: menengah tinggi {menengah=0, menengah tinggi=2}
|   Kondisi Cuaca = cerah: menengah tinggi {menengah=0, menengah tinggi=3}

```

Gambar 5. Rule

Akurasi merupakan suatu nilai dari seberapa baik model mengkorelasikan antara hasil dengan atribut dalam data yang telah disediakan.

☒ Table View ☐ Plot View

accuracy: 100.00%

	true menengah	true menengah tinggi	class precision
pred. menengah	0	0	0.00%
pred. menengah tinggi	0	2	100.00%
class recall	0.00%	100.00%	

Gambar 6. Tampilan Accuration

Berdasarkan gambar yang telah dijelaskan di atas, maka dapat diketahui bahwa pengujian data yang dilakukan dengan menggunakan *apply model* dan % *Performance* diperoleh nilai akurasi 100% dapat dikategorikan model algoritma *C4.5* ini adalah *excellent*.

Performance

Description

Annotations

PerformanceVector

PerformanceVector:
accuracy: 100.00%

ConfusionMatrix:
True: menengah menengah tinggi
menengah: 0 0
menengah tinggi: 0 2
precision: 100.00% (positive class: menengah tinggi)

ConfusionMatrix:
True: menengah menengah tinggi
menengah: 0 0
menengah tinggi: 0 2
recall: 100.00% (positive class: menengah tinggi)

ConfusionMatrix:
True: menengah menengah tinggi
menengah: 0 0
menengah tinggi: 0 2
AUC (optimistic): unknown (positive class: menengah tinggi)
AUC: 0.500 (positive class: menengah tinggi)
AUC (pessimistic): unknown (positive class: menengah tinggi)

Gambar 7. Tampilan Accuration

4. Kesimpulan

Berdasarkan hasil penelitian diatas maka dapat disimpulkan bahwa:

- a) Kenaikan rata-rata hasil ikan tangkapan Dapat disebabkan oleh faktor Jumlah nelayan, Kondisi Cuaca, Kondisi Air, Lebar Jaring dan jumlah trip berlayar. Kurangnya koordinasi dari pihak terkait membuat perbedaan dari data yang ada. Dari data yang didapat di simpulkan bahwa faktor terbesar penyebab terjadinya kenaikan rata-rata ikan tangkapan adalah karena jumlah nelayan yang kurang pada saat berlayar menangkap ikan perharinya.
- b) Algoritma C4.5 merupakan algoritma *Decision tree* (Pohon Keputusan) yang mempunyai input berupa training sampel dan sampel. Training sampel berupa data contoh yang digunakan dalam membangun sebuah pohon keputusan yang telah diuji kebenarannya. Sedangkan sampel merupakan *field-field* data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data.
- c) Data yang digunakan diperoleh dari situs resmi BPS Simlaungun yaitu data hasil ikan tangkapan dan faktornya. Setelah melakukan *cleaning data*, data tersebut diolah dengan menggunakan algoritma C4.5 dan dilakukan pengujian data menggunakan *software RapidMiner*. *RapidMiner* menunjukkan hasil yang sama yaitu atribut Jumlah nelayan dengan tingkat akurasi 100%.

Daftar Pustaka

- [1] H. Rofika, "Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate Menggunakan Algoritma K-Means," *Jiko (Jurnal Inform. Dan Komputer)*, Vol. 2, No. 1, Pp. 26–33, 2019, Doi: 10.33387/Jiko.V2i1.1053.
- [2] A. Chandra, "Penerapan Data Mining Menggunakan Pohon Keputusan," Vol. 6, No. 3, Pp. 1–6, 2017.
- [3] E. Elisa, "Analisa Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi Pt.Arupadhatu Adisesanti," *J. Online Inform.*, Vol. 2, No. 1, P. 36, 2017, Doi: 10.15575/Join.V2i1.71.
- [4] N. Azwanti, "Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, Vol. 13, No. 1, P. 33, 2018, Doi: 10.30872/Jim.V13i1.629.
- [5] M. F. Arifin And D. Fitriana, "Penerapan Algoritma Klasifikasi C4.5 Dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : Pt Atria Artha Persada," *Incomtech*, Vol. 8, No. 2, Pp. 87–102, 2018, Doi: 10.22441/Incomtech.V8i1.2198.
- [6] J. Jurnal And S. Informasi, "Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4 . 5 Stmik Royal , Ksian," Vol. 2, 2016.
- [7] S. Kasus And R. S. U. Provinsi, "Klasifikasi Penyakit Hipertensi Classification Of Hypertension Disease Using C4 . 5 Algorithm," 2017.
- [8] S. Kasus, U. Dehasen, S. Haryati, A. Sudarsono, And E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4 . 5," *J. Media Infotama*, Vol. 11, No. 2, Pp. 130–138, 2015.