

Prediksi Kenaikan Rata-Rata Volume Perikanan Tangkap Dengan Teknik *Data Mining*

Anik Andriani

AMIK BSI Yogyakarta

E-Mail: anik.aai@bsi.ac.id

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 Decision Tree, Naive Bayes, Neural Network, dan Support Vector Machine. Hasil klasifikasi diuji dengan confusion matrix dan kurva ROC untuk mengetahui tingkat performance masing-masing algoritma 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*, *Naive Bayes*, *Neural Network*

1. PENDAHULUAN

Pertumbuhan penduduk di Indonesia yaitu sebesar 1,8% pertahun dan saat ini jumlahnya mencapai lebih dari 250 juta jiwa. Hal tersebut mendorong meningkatnya jumlah permintaan terhadap kebutuhan sumber daya dan jasa lingkungan. Untuk mencukupi kebutuhan pangan yang merupakan kebutuhan pokok, ketersediaan sumber daya alam yang berada di darat semakin berkurang dan tidak mencukupi, sehingga perlu penambahan pemanfaatan sumber daya alam yang berada di laut dan pesisir [1]. Salah satu cara dalam pemanfaatan sumber daya alam di laut dan pesisir yaitu dengan perikanan tangkap. Kondisi perikanan tangkap di Indonesia saat ini belum optimal, beberapa daerah mengalami kemunduran produktivitas hasil perikanan tangkap yang berdampak pada minimnya kontribusi terhadap peningkatan produktivitas perikanan tangkap [2]. Sedangkan volume produksi perikanan secara keseluruhan mengalami sedikit peningkatan dari tahun ke tahun yaitu sekitar

3,2% [3]. Evaluasi kenaikan volume produksi perikanan dapat dilakukan dengan mengevaluasi faktor-faktor yang mempengaruhinya. Penelitian ini bertujuan membuat klasifikasi data untuk mengetahui faktor yang berpengaruh terhadap kenaikan volume perikanan tangkap dengan menerapkan teknik data mining. Data mining adalah analisis pengamatan dataset yang besar untuk menemukan hubungan yang tidak terduga dari dataset tersebut dan untuk meringkas data dengan cara yang dapat dimengerti dan dapat dimanfaatkan oleh pemilik data tersebut [4].

Pada penelitian sebelumnya [5] dibangun model prediksi perkiraan permintaan pelanggan terhadap produksi ikan dengan algoritma Back Propagation Neural Network. Hasil penelitian menunjukkan tingkat kesuksesan algoritma Back Propagation Neural Network dalam membuat prediksi permintaan produksi ikan pada data training rata-rata sebesar 100%, sedangkan tingkat kesuksesan pada data testing rata-rata sebesar 64,3%.

Berdasarkan latar belakang diatas maka rumusan masalah dalam penelitian ini dapat diuraikan dalam beberapa Research Question berikut:

1. Bagaimanakah hasil penerapan algoritma klasifikasi dalam membuat prediksi naik atau turunnya volume produksi perikanan tangkap?
2. Bagaimanakah tingkat performance algoritma dalam mengklasifikasi data?
3. Algoritma klasifikasi manakah yang paling tinggi akurasi?

2. TINJAUAN PUSTAKA

2.1. Perikanan Tangkap

Perikanan tangkap merupakan kegiatan ekonomi yang dilakukan mencakup kegiatan penangkapan atau pengumpulan hewan dan tanaman air yang hidup di laut atau pada perairan umum secara bebas [6]. Usaha pengembangan perikanan tangkap memerlukan dukungan teknologi, salah satunya adalah teknologi alat tangkap dengan tingkat selektifitas yang tinggi yang dapat dioperasikan untuk eksploitasi ikan laut dalam [7]. Jenis alat tangkap yang dominan digunakan adalah jaring insang, jaring udang, pukat cincin [8]. Pada penelitian ini jenis alat tangkap yang dijadikan atribut antara lain: kapal, kapal motor, perahu tanpa motor, perahu motor tempel, pukat tarik, pukat kantong, pukat cincin, jaring angkat, dan pancing.

2.2. Data Mining

Data mining adalah ekstraksi atau penambangan pengetahuan dari data yang besar [9]. Penambangan data dengan data mining merupakan proses untuk menemukan pola dalam sebuah dataset dengan proses yang otomatis maupun semi-otomatis. Pola tersebut diidentifikasi, divalidasi, dan digunakan untuk membuat prediksi [10]. Data mining dapat dicapai dengan beberapa teknik antara lain Association, Classification, Clustering, Prediction, Sequential Pattern, dan Similiar Time Sequences [11]. Teknik klasifikasi merupakan proses untuk menempatkan objek dalam dataset kedalam satu set kategori berdasarkan sifat dari masing-masing objek [12]. Contoh klasifikasi adalah memprediksi apakah pelanggan membeli atau tidak, memprediksi sebuah pengeboran menemukan minyak atau tidak

[13]. Algoritma yang akan digunakan untuk klasifikasi yaitu *Decision Tree*, *Naive Bayes*, *Neural Network*, dan *Support Vector Machine*.

2.2.1. Decision Tree

Algoritma ini mengubah fakta yang sangat besar menjadi representasi aturan dalam bentuk pohon keputusan [14]. *Decision tree* mendeteksi karakteristik data yang membuat prediksi data dimasa mendatang yang berkaitan dengan atribut pada kelas tertentu [15]. Salah satu algoritma dalam *Decision Tree* adalah Algoritma C4.5 dimana pembelajarannya memetakan satu set data yang hasilnya dapat diterapkan untuk kasus lainnya [16]. Cara kerjanya adalah mencari atribut yang dijadikan akar dengan nilai gain tertinggi. Dalam perhitungan nilai gain dihitung terlebih dahulu nilai entropy dengan rumus berikut:

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

S adalah Himpunan kasus. n adalah jumlah partisi S. P_i adalah proporsi S_i terhadap S.

Untuk perhitungan nilai gain dengan rumus berikut:

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

S adalah Himpunan Kasus. A adalah Fitur. n adalah jumlah partisi atribut A. $|S_i|$ adalah Proporsi S_i terhadap S. $|S|$ adalah jumlah kasus dalam S.

Perhitungan nilai gain diulangi terus hingga semua record terpartisi.

2.2.2. Naive Bayes

Algoritma Naive Bayes mudah dibangun dan tidak memerlukan skema estimasi parameter berulang yang rumit [16]. Berikut rumus teorema Bayes:

$$P(x/y) = \frac{P(y/x) P(x)}{P(y)} \quad (3)$$

y adalah kelas yang belum diketahui. x adalah hipotesis data y merupakan suatu kelas spesifik. $P(x|y)$ adalah probabilitas hipotesis x berdasar kondisi y (*posteriori probability*). $P(x)$ adalah probabilitas hipotesis x (*prior probability*). $P(y|x)$ adalah probabilitas y berdasarkan kondisi pada hipotesis x. $P(y)$ adalah probabilitas dari y.

2.2.3. Neural Network

Neural Networks (NNS) mempunyai karakteristik fundamental yang mampu untuk belajar dari training data “dengan atau tanpa guru” [12]. Sekitar 95% aplikasi bisnis yang menggunakan Neural Network memakai algoritma Multilayer Perceptron (MLP) [17]. Penemuan algoritma backpropagation untuk MLP merupakan metode sistematis untuk training sehingga lebih efisien [18].

2.2.4. Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu algoritma yang kuat dan akurat dari algoritma-algoritma dalam data mining [19]. SVM Classifiers membuat sebuah maximum-margin hyperplane yang terletak di ruang input ditransformasikan dan dibagi kedalam contoh class [18]. Pendekatan SVM cocok untuk kasus nonlinier [20].

2.3. Evaluasi

Ukuran tingkat performance hasil klasifikasi dapat menggunakan akurasi. Akurasi dalam klasifikasi adalah persentase ketepatan record data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi [21]. Akurasi merupakan rincian dari kinerja classifier (hasil untuk kasus yang tidak diketahui) mengklasifikasikan kejadian X dengan benar kedalam class X [22]. Metode yang digunakan untuk pengujian antara lain:

2.3.1. Confussion Matrix

Metode confusion matrix menggunakan tabel matriks berikut

Tabel 1. Model Confussion Matrix [21]

Correct Classification	Classified as	
	+	-
+	True positives	False negatives
-	False positives	True negatives

2.3.2. Kurva ROC (Receiver Operating Characteristic)

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual.

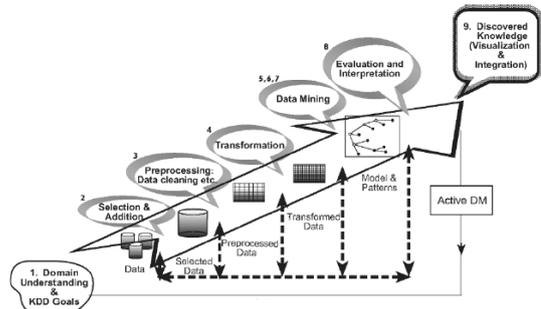
ROC mengekspresikan confusion matrix[17]. The area under curve (AUC) dihitung untuk mengukur perbedaan performansi metode yang digunakan [23]. Nilai akurasi dari kurva ROC dikategorikan seperti Tabel 2.

Tabel 2. Kategori hasil kurva ROC [21]

Nilai akurasi	Kategori
0,90-1,00	Excellent Classification
0,80-0,90	Good Classification
0,70-0,80	Fair Classification
0,60-0,70	Poor Classification
0,50-0,60	Failure

3. METODE PENELITIAN

Tahapan penelitian dalam klasifikasi ini mengadopsi sembilan langkah dalam Knowledge Discovery in Databases (KDD) seperti yang ditunjukkan Gambar 1.



Gambar 1. Knowledge Discovery in Databases (KDD) [18]

Sembilan langkah dalam KDD antara lain:

1. *Developing an understanding of the application domain*, merupakan tahap persiapan untuk menentukan langkah-langkah dalam penelitian
2. *Selection and creating a dataset on which discovery will be performed*, menyeleksi dan membuat data penelitian. Dalam penelitian ini data yang digunakan untuk penelitian adalah data statistik volume rata-rata produksi perikanan tangkap Kementerian Kelautan dan Perikanan. Dataset yang tersedia sebanyak 132 data dengan jumlah atribut 13 atribut dan 1 class.
3. *Preprocessing and cleansing*, tahap untuk meningkatkan kehandalan data dengan membuang data yang tidak lengkap (missing value) dan atau tidak benar (noise). Pada tahap ini diperoleh data sebanyak 130 data.

4. *Data transformation*, tahap meningkatkan generasi data yang lebih baik yaitu dengan mentransformasi data kedalam bentuk kategori. Selain itu pada tahap ini data dibagi menjadi dua bagian yaitu data training (80%) dan data testing (20%) dengan teknik Systematic Random Sampling. Hasilnya diperoleh data training sebanyak 104 data dan data testing sebanyak 26 data.
5. *Choosing the appropriate data mining task*, tahap menentukan teknik data mining yang akan dipakai yaitu klasifikasi.
6. *Choosing the data mining algorithm*, tahap menentukan algoritma yang akan dipakai untuk klasifikasi yaitu Decision Tree, Naive Bayes, Neural Network, dan SVM.
7. *Employing the data mining algorithm*, tahap pengolahan data dengan algoritma yang telah dipilih.
8. *Evaluation*, tahap mengevaluasi hasil klasifikasi untuk mengukur tingkat performance hasil klasifikasi dengan menggunakan Confusion Matrix dan Kurva ROC.
9. *Using the discovered knowledge*, tahap penerapan pengetahuan yang diperoleh dari hasil klasifikasi dengan data mining.

4. HASIL DAN PEMBAHASAN

Hasil klasifikasi diuji dengan confusion matrix untuk mengetahui tingkat performance algoritma dalam membuat klasifikasi yang ditunjukkan dengan persentase nilai akurasi. Pengujian dilakukan dengan menguji hasil klasifikasi dengan data training yang digunakan untuk membuat klasifikasi itu sendiri dan dengan data testing (data baru). Berikut hasil pengujian klasifikasi data.

Tabel 5. Nilai akurasi terhadap data training dengan confusion matrix

Decision Tree	Naive Bayes	SVM	Neural Network
87,50%	74,04%	74,04%	85,58%

Tabel 6. Nilai akurasi terhadap data testing dengan confusion matrix

Decision Tree	Naive Bayes	SVM	Neural Network
53,85%	61,54%	61,54%	65,38%

Tabel 5 menunjukkan nilai akurasi tertinggi dari hasil klasifikasi dengan confusion matrix yaitu 87,50% dengan algoritma Decision Tree. Sedangkan Tabel 6 menunjukkan pengujian dengan data testing yaitu nilai akurasi tertinggi sebesar 65,38% dengan algoritma Neural Network. Pengujian dengan kurva ROC/AUC ditunjukkan pada Tabel berikut.

Tabel 7. Nilai akurasi terhadap data training dengan kurva ROC

Decision Tree	Naive Bayes	SVM	Neural Network
0,977	0,859	0,769	0,923

Tabel 8. Nilai akurasi terhadap data testing dengan kurva ROC

Decision Tree	Naive Bayes	SVM	Neural Network
0,667	0,614	0,686	0,712

Tabel 7 menunjukkan nilai akurasi hasil klasifikasi terhadap data training dengan menggunakan kurva ROC yaitu algoritma Decision Tree memiliki nilai tertinggi sebesar 0,977. Sedangkan Tabel 8 menunjukkan nilai akurasi hasil klasifikasi terhadap data testing dengan kurva ROC yaitu algoritma Neural Network sebesar 0,712.

Berdasarkan perbandingan nilai akurasi yang diperoleh dari pengujian hasil klasifikasi terhadap data training, menunjukkan algoritma Decision Tree lebih baik dibanding dengan algoritma yang lain. Nilai 0,977 yang ditunjukkan pada kurva ROC menunjukkan bahwa klasifikasi masuk kategori *excellent classification*. Tetapi pada saat hasil klasifikasi diuji dengan data testing menunjukkan algoritma Neural Network lebih baik dibanding algoritma yang lain, dengan nilai 0,712 yang masuk kategori *fair classification*.

5. KESIMPULAN

Penerapan algoritma Decision Tree, Naive Bayes, SVM, dan Neural Network pada kasus prediksi kenaikan volume rata-rata perikanan tangkap cukup baik. Algoritma Neural Network mempunyai nilai akurasi tertinggi dalam membuat klasifikasi pada kasus tersebut. Tetapi hasil klasifikasi masih tergolong kategori *fair classification*.

Penelitian selanjutnya dapat menerapkan attribute selection untuk menyeleksi atribut yang mempunyai nilai korelasi tinggi terhadap output class sehingga diharapkan dapat meningkatkan nilai akurasi hasil klasifikasi.

6. DAFTAR PUSTAKA

- [1] D. G. Bengen, "Ekosistem dan Sumberdaya Pesisir dan Laut serta Pengelolaan Secara Terpadu dan Berkelanjutan," in *Pelatihan Pengelolaan Wilayah Pesisir Terpadu*, Bogor, 2001, pp. 28-55.
- [2] R. Noviyanti, "Kondisi Perikanan Tangkap di Wilayah Pengelolaan Perikanan (WPP) Indonesia," in Seminar FMIPA-UT 2011, Jakarta, 2011
- [3] A. Samosir, "Sektor Perikanan: PNBPN yang Terabaikan," 2013.
- [4] D. T. Larose, *Data Mining Methods and Models*. New Jersey: Wiley, 2006.
- [5] C. Yao Lo, C. I Hou, and Y. Yun Pai, "An Intelligent Demand Forecasting Model with Back Propagation Neural Network for Fish Product," *Journal of Applied Sciences, Engineering and Technology*, vol. III, no. 5, pp. 447-455, May 2011.
- [6] D. Monintja and R. Yusfiandayani, "Pemanfaatan Sumber Daya Pesisir dalam Bidang Perikanan Tangkap," in *Pelatihan Pengelolaan Wilayah Pesisir Terpadu*, Bogor, 2001, pp. 56-65.
- [7] H. E. Irianto and I. Soesilo, "Dukungan Teknologi Penyediaan Produk Perikanan," in *Seminar Nasional Hari Pangan Sedunia*, Bogor, 2007.
- [8] T. Resmiati, S. Diana, and S. Astuty, "Komposisi Jenis Alat Tangkap yang Beroperasi di Perairan Teluk Banten, Serang," *Universitas Padjadjaran Laporan Penelitian*, 2002.
- [9] S. Dua and X. Du, *Data Mining and Machine Learning in Cybersecurity*. United States of Amerika: CRC Press, 2011.
- [10] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Techniques 3rd Edition*. Burlington: Elsevier, 2011.
- [11] D. L. Olson and D. Delen, *Advances Data Mining Techniques*. Berlin, German: Springer, 2008.
- [12] F. Gorunescu, *Data Mining Concepts, Models and Techniques*. Berlin: Springer, 2011.
- [13] G. J. Myatt, *Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining*. New Jersey, Canada: John Wiley & Sons, 2007.
- [14] Kusriani and E. T. Lutfi, *Algoritma Data Mining*, T. A. Prabawati, Ed. Yogyakarta, Indonesia: Andi Offset, 2009.
- [15] I. Ntoutsi, N. Pelekis, and Y. Theodoridis, "Pattern Comparison Data Mining: A Survey," in *Research and Trends in Data Mining Technologies and Applications*, K. Klinger, et al., Eds. USA: IDEA Group Publishing, 2007, ch. IV, pp. 86-120.
- [16] X. Wu and V. Kumar, *The Top Ten Algorithms in Data Mining*. New York: CRC Press, 2009.
- [17] C. Verrellis, *Business Intelligence, Data Mining and Optimization for Decision Making*. United Kingdom: John Wiley & Sons, 2009.
- [18] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*. New York: Springer, 2010.
- [19] H. Xue, Q. Yang, and S. C. Chen, "SVM: Support Vector Machines," in *The Top Ten Algorithms in Data Mining*, W. Xindong and V. Kumar, Eds. Boca Raton, USA: CRC Press, 2009, ch. 3, pp. 37-58.
- [20] Z. Markov and D. T. Larose, *Data Mining The Web*. New Jersey, Canada: John Wiley and Sons, 2007.
- [21] J. Han and M. Kamber, *Data Mining Concepts and Techniques 2nd edition*. USA: Elsevier, 2006.
- [22] M. Bramer, *Principles of Data Mining*. London: Springer, 2007.
- [23] S.-H. Liao, "Expert System Methodologies and Application-a decade review from 1995 to 2004," *Elsevier*, pp. 1-11, 2004.