

HYBRIDIZATION OF THE NAIVE BAYES CLASSIFICATION METHOD IN THE FRESHWATER FISH SEED SELLER CLASSIFICATION MODEL

M. Hafidz Ariansyah^{*1}, Esmi Nur Fitri², Sri Winarno³, Asih Rohmani⁴, Fikri Budiman⁵, Junta Zeniarja⁶
Edi Sugiarto⁷

^{1,2,3,4}Departemen Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Indonesia
^{5,6,7}Departemen Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Indonesia
Email: ¹11220190614@mhs.dinus.ac.id, ²11220190614@mhs.dinus.ac.id, ³sri.winarno@dsn.dinus.ac.id,
⁴aseharsoyo@dsn.dinus.ac.id, ⁵fikri.budiman@dsn.ac.id, ⁶junta@dsn.ac.id, ⁷edi.sugiarto@dsn.ac.id

(Naskah masuk: 29 November 2022, Revisi: 8 Desember 2022, Diterbitkan: 23 Maret 2023)

Abstract

Freshwater fish seed sellers play several roles in the supply chain process in the freshwater fish farming business. The role of the seller of freshwater fish seeds in this process is to distribute fish seeds which are one of the upstream sources in the supply chain process. Freshwater fish cultivators must select competent freshwater fish seed sellers so the supply chain process can run well. A large number of freshwater fish seed sellers in the market remind freshwater fish cultivators to choose the quality of the freshwater fish seed seller in terms of seed quality, low prices, shipping that can reach many areas, ergonomic packaging, and others. This study proposes Hybrid Naïve Bayes Classifiers (HNBCs) as a machine learning method for classification. This study aimed to compare the seed seller classification method in which the appropriate pattern of seed seller was identified by hybridization of Naïve Bayes Classifiers (NBCs), and then the researchers conducted performance appraisal and evaluation. The results are beneficial for freshwater fish cultivators and researchers which will enable them to formulate their plans according to the predicted results. The proposed method has produced significant results by achieving a training data accuracy of 82.61% and the testing data accuracy of 73.91%.

Keywords: *Classification, Fish Farming, Fish Seed Seller, Hybrid Naive Bayes Classifiers, Machine Learning*

HIBRIDISASI METODE KLASIFIKASI NAIVE BAYES DALAM MODEL KLASIFIKASI PENJUAL BENIH IKAN AIR TAWAR

Abstrak

Penjual benih ikan air tawar memainkan sejumlah peran dalam proses rantai pasokan dalam proses bisnis pembudidayaan ikan air tawar. Peran penjual benih ikan air tawar dalam proses tersebut adalah melakukan distribusi benih ikan yang menjadi salah satu hulu dalam proses rantai pasok. Para pembudidaya ikan air tawar harus memilih penjual benih ikan air tawar yang kompeten agar proses rantai pasok dapat berjalan dengan baik. Banyaknya penjual benih ikan air tawar di pasar mengingatkan pembudidaya ikan air tawar agar memilih penjual benih ikan air tawar yang berkualitas, baik dari segi kualitas benih, harga yang murah, pengiriman yang dapat menjangkau banyak area, pengemasan yang ergonomis, dan lain sebagainya. Penelitian ini mengusulkan *Hybrid Naïve Bayes Classifiers (HNBCs)* sebagai metode pembelajaran mesin untuk klasifikasi. Sasaran dari penelitian ini adalah untuk membandingkan metode klasifikasi penjual benih ikan air tawar dimana pola penjual benih ikan air tawar yang sesuai diidentifikasi dengan hibridisasi *Naïve Bayes Classifiers (NBCs)*, dan selanjutnya peneliti melakukan penilaian kinerja dan evaluasi. Hasilnya sangat penting bagi pembudidaya ikan air tawar dan peneliti yang akan memungkinkan mereka untuk merumuskan rencana mereka sesuai dengan hasil prediksi. Metode yang diusulkan telah menghasilkan hasil yang cukup signifikan dengan mencapai akurasi data training sebesar 82,61% dan akurasi testing sebesar 73,91%.

Kata kunci: *budidaya ikan, hybrid naive bayes classifiers, klasifikasi, machine learning, penjual benih ikan*

1. PENDAHULUAN

Budidaya ikan air tawar merupakan penyumbang produksi ikan terbesar di Indonesia. Perikanan darat memberikan kontribusi terhadap

sektor perikanan sebesar 53,7%, lebih tinggi dari perikanan laut sebesar 46,3% [1]. Sebagai contoh, menurut data Badan Pusat Statistik, produksi ikan nila yang dibudidayakan pada tahun 2020 mencapai 364.747 ton, dan produksi lele sebesar 347.511 ton

[1]. Angka tersebut merupakan angka yang besar meskipun angka di atas hanya merupakan hasil perhitungan sementara dan akan terus bertambah melalui proses perhitungan data lainnya. Namun, angka tersebut belum dapat memenuhi permintaan konsumsi ikan air tawar di pasar dalam negeri maupun luar negeri sehingga pembudidaya harus meningkatkan produksi ikan untuk memenuhi permintaan pasar. Karena permintaan pasar yang besar, pembudidaya membutuhkan banyak benih ikan. pembudidaya biasanya membeli benih ikan dari penjual benih ikan. Penjual benih ikan adalah mitra usaha yang menjamin ketersediaan produk atau bahan baku yang dibutuhkan oleh suatu instansi atau perorangan [2]. Saat ini banyak pembudidaya yang melakukan seleksi terhadap penjual benih ikannya. Pemilihan penjual benih ikan adalah masalah multi-kriteria penting yang dapat mempengaruhi bisnis. Isu selanjutnya adalah keputusan penjual benih ikan reguler yang berdampak besar pada aspek ekonomi, sosial, dan lingkungan. Pemilihan penjual benih ikan merupakan tahapan proses pembelian industri yang biasanya dilakukan dengan mempertimbangkan banyak atribut penjual benih ikan dan kepentingannya [3]. Variabel dalam pemilihan dapat meliputi kualitas produk dan layanan, reputasi, pengiriman tepat waktu, perilaku perusahaan yang etis, komunikasi yang jujur, dan harga bersaing. Oleh karena itu, pembudidaya dan pembeli dapat memilih penjual benih ikan tunggal maupun beberapa penjual benih ikan lain yang dapat menunjang perolehan benih [4].

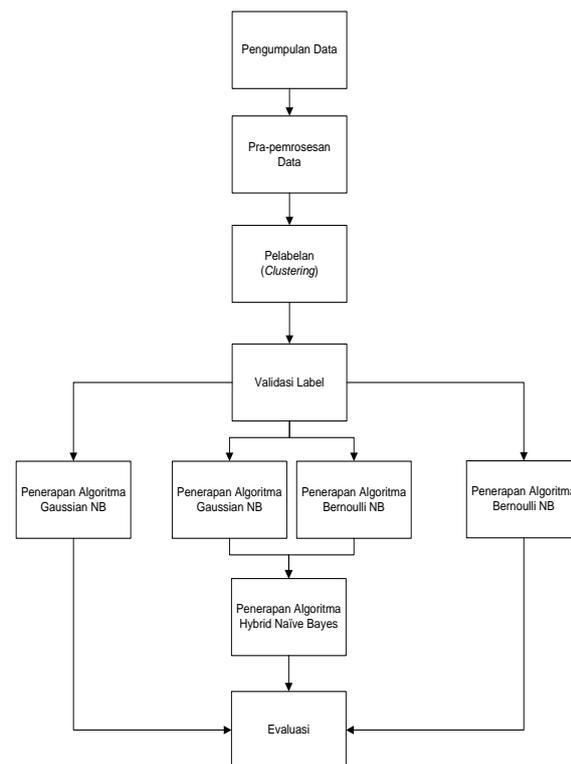
Berbagai metode dapat diadopsi dalam mengevaluasi proses pengambilan keputusan pemilihan penjual benih ikan, termasuk metode klasifikasi dalam data mining. Teknik klasifikasi memiliki banyak aplikasi di berbagai bidang ilmu. Data pelatihan digunakan untuk membuat model kelas guna mendapatkan label kelas untuk sampel baru dalam model klasifikasi. Output dari model klasifikasi dapat berupa varian seperti pada *Classifier Decision Tree* atau kontinu seperti pada *Classifier Naive Bayes* [5]. Salah satu konferensi internasional terbesar tentang data mining, IEEE International Conference on Data Mining (ICDM) menominasikan sepuluh besar algoritma data mining, yaitu *C4.5*, *K-Means*, *Support Vector Machines (SVM)*, *Apriori*, *Expectation Maximization Algorithm (EM)*, *Page Rank*, *AdaBoost*, *K-Nearest Neighbors (KNN)*, *Naive Bayes*, *Classification and Regression Tree (CART)* [6].

Penelitian ini memanfaatkan algoritma *Gaussian Naive Bayes*, *Bernoulli Naive Bayes*, dan *Hybrid Naive Bayes* yang didapatkan dari penggabungan algoritma sebelumnya untuk mengklasifikasi penjual benih ikan sehingga pembudidaya bisa mendapatkan rekomendasi penjual benih ikan berdasarkan variabel yang ada. Penelitian ini akan mencari nilai akurasi terbaik dari

ketiga algoritma yang digunakan agar prediksi atau pelabelan dapat dilakukan dengan baik. Tujuan dari penelitian ini adalah untuk membandingkan akurasi pola label rekomendasi dan tidak rekomendasi sehingga dapat menjadi dasar pengambilan keputusan bagi pembudidaya ikan air tawar.

2. METODE PENELITIAN

Metode yang dilakukan di dalam penelitian secara garis besar ditunjukkan pada Gambar 1.



Gambar 1. Gambaran Penelitian

Penelitian ini memanfaatkan metode *data mining* untuk menemukan pola atau kemungkinan dari berbagai kumpulan data. Singkatnya, ini mempelajari model untuk melakukan klasifikasi terhadap rekomendasi penjual benih ikan air tawar. Pengelompokan dari data yang ada menggunakan *Gaussian Mixtures Clustering*; melatih model *Mixed Gaussian-Bernoulli Naive Bayes (Hybrid Naive Bayes)*, dengan validasi silang 10 kali lipat; dan, melakukan kebingungan analisis matriks untuk menguji akurasi kinerja model. Tahapan dalam penelitian ini terdiri dari 6 proses, yaitu pengumpulan data, pra-pemrosesan data, pelabelan (*clustering*), validasi label, penerapan algoritma, dan evaluasi seperti terdapat pada Gambar 1.

2.1. Pemilihan Penjual Benih Ikan

Pemilihan penjual benih ikan merupakan kegiatan yang strategis, khususnya pada saat penjual benih ikan menyediakan benih ikan yang digunakan pada awal proses bisnis dalam suatu rantai pasok produksi. Kriteria pemilihan merupakan salah satu

hal yang paling penting saat memilih *supplier* [7]. Penelitian [8] menunjukkan bahwa kriteria pemilihan pemasok dapat sangat bervariasi. Tabel 1 menunjukkan bahwa Dickson mengidentifikasi 22 kriteria [8]. Angka di kolom hasil menunjukkan pentingnya setiap kriteria, berdasarkan kumpulan tanggapan dari survei terhadap 170 manajer pembelian di Amerika Serikat. Responden diminta untuk memilih angka dari 0 sampai 4 pada skala Likert, dengan 0 sangat tidak penting dan 4 sangat penting.

Tabel 1. Kriteria pemilihan pemasok Dickson

No	Kriteria	Nilai
1	Kualitas	3.5
2	Pengiriman	3.4
3	Riwayat Performa	3.0
4	Jaminan	2.8
5	Harga	2.8
6	Kemampuan Teknis	2.8
7	Posisi Finansial	2.5
8	Pematuhan Prosedur	2.5
9	Sistem Komunikasi	2.5
10	Reputasi dan Posisi di Industri	2.4
11	Keinginan untuk bisnis	2.4
12	Manajemen	2.3
13	Kontrol Operasi	2.2
14	Servis Perbaikan	2.2
15	Sikap	2.1
16	Kesan	2.1
17	Kemampuan Pengepakan	2.0
18	Hubungan Tenaga Kerja	2.0
19	Lokasi Geografis	1.9
20	Jumlah Bisnis Sebelumnya	1.6
21	Alat Bantu Pelatihan	1.5
22	Pengaturan Timbal Balik	0.6

2.2. Pengumpulan Data

Data merupakan kumpulan fakta mentah yang tidak memiliki arti sehingga perlu diolah menjadi informasi agar memiliki nilai pengetahuan [9]. Data penjual ikan air tawar merupakan data yang digunakan dalam memperoleh dari *e-marketplace* dan diberi label sebelumnya. Data ini memiliki 11 atribut yaitu Bibit, uk2_3, uk3_4, uk4_6, uk7_9, uk9_12, Remaja, Pengiriman, Kemasan, Diskon, Garansi, dan Jumlah Penilaian serta memiliki 92 data. Atribut dalam dataset ini mengacu pada penelitian sebelumnya yang dilakukan oleh Dickson [8]. Deskripsi atribut dan dataset (lihat Tabel 2) yang digunakan adalah sebagai berikut:

1. Bibit - Jenis bibit yang dijual
2. uk2_3 - Harga bibit ukuran 2 cm - 3 cm (dalam rupiah)
3. uk3_4 - Harga bibit ukuran 3 cm - 4 cm (dalam rupiah)
4. uk4_6 - Harga bibit ukuran 4 cm - 6 cm (dalam rupiah)
5. uk7_9 - Harga bibit ukuran 7 cm - 9 cm (dalam rupiah)
6. uk9_12 - Harga benih ukuran 9 cm - 12 cm (dalam rupiah)
7. Remaja - Harga bibit lebih dari 12 cm (dalam rupiah)

8. Pengiriman - Jarak pengiriman (jauh/dekat)
9. Diskon - Jumlah diskon
10. Garansi - Ya / Tidak
11. Jumlah Penilaian - Rendah / Sedang / Tinggi

Tabel 2. Dataset

	Clarias	Clarias	...	Bramidae
Bibit				
uk2_3	150	150	...	450
uk3_4	150	150	...	750
uk4_6	220	200	...	1000
uk7_9	450	450	...	1200
uk9_12	600	500	...	3200
Remaja	1100	1100	...	7000
Pengiriman	Jauh	Jauh	...	Jauh
Diskon	1	1	...	1
Jaminan	Ya	Ya	...	Tidak
Jumlah Penilaian	Rendah	Sedang		Rendah

2.3. Pra-pemrosesan Data

Pra-pemrosesan data merupakan salah satu teknik yang digunakan sebelum tahap analisis data, dimana yang perlu dilakukan adalah memeriksa apakah data tersebut memiliki *missing value*, duplikasi data, atau inkonsistensi data [10]. Untuk mendapatkan data yang berkualitas tinggi, peneliti melakukan berbagai *preprocessing* data dalam menghilangkan data dan mengidentifikasi proses, yaitu melalui data yang tidak konsisten, data yang tidak relevan (*noise/outlier*), dan data yang tidak lengkap [11]. Selain itu peneliti melakukan proses seperti pengecekan data yang tidak konsisten, melakukan perbaikan data, dan menghilangkan duplikasi data [12]. Dari data yang terkumpul akan dianalisis apabila terdapat data yang tidak konsisten atau tidak relevan yang mengganggu pola pengambilan data algoritma yang akan dibentuk.

2.4. Proses Pelabelan dengan *Clustering*

Setelah mengumpulkan dan membersihkan data dalam tahap pra-pemrosesan, data kemudian diproses menggunakan *Machine Learning* dengan metode yang disebut *Gaussian Mixtures Clustering* untuk mengidentifikasi yang mana di antara penjual benih ikan dapat dicap sebagai Rekomendasi dan Tidak Rekomendasi. Pengelompokan sering melibatkan kategori data. Dengan demikian, pendekatan konvensional dari mengubah data kategorikal menjadi data numerik tidak tentu menghasilkan hasil yang berarti. Untuk fase ini, metode *clustering* yang digunakan adalah *Gaussian Mixtures Clustering*. *Gaussian Mixtures Clustering* adalah teknik analisis data *unsupervised* yang digunakan untuk mengklasifikasikan data yang sama ke dalam kelompok homogen, itu bekerja secara khusus dengan baik pada data kategorikal [13][14]. Dengan menggunakan analisis *Gaussian Mixtures Clustering* mengelompokkan sejumlah besar data ke dalam subset yang disebut *cluster*. Setiap *cluster* adalah kumpulan data objek (penjual dalam hal ini) yang mirip dengan satu lain ditempatkan dalam *cluster* yang sama tetapi berbeda dengan objek di *cluster* lain. Data dipisahkan menjadi 2 *cluster* untuk

mengidentifikasi penjual benih ikan yang dianggap rekomendasi untuk dibeli, dan tidak rekomendasi untuk dibeli. Setelah fitur masukan disusun dan data keluaran diidentifikasi, kedua data tersebut digabungkan dalam satu dataset.

2.5. Algoritma Naïve Bayes

Naïve Bayes Classifier adalah pengklasifikasi pembelajaran mesin terawasi probabilistik yang menggunakan aturan *Bayes* yang mencakup semua fitur dalam data telah mengasumsikan fitur independensi. Ini berarti bahwa tidak ada hubungan antara yang berbeda nilai fitur. Selain itu, *Naïve Bayes Classifier* mewakili pengawasan yang paling banyak digunakan pengklasifikasi pembelajaran mesin dalam domain penambangan teks dan aplikasi penambangan data karena dicirikan dengan kesederhanaan dan efektivitas [15]. *Naïve Bayes Classifier* memiliki empat model, *Gaussian Naïve Bayes*, *Multinomial Naïve Bayes*, *Bernoulli Naïve Bayes*, dan *Categorical Naïve Bayes*. Hibridisasi algoritma model *classifier* ini dengan jumlah parameter yang berbeda dan fitur mengarah pada pencapaian optimalisasi [16][17]. Ini berarti tidak ada hubungan antara nilai fitur yang berbeda [18]. Studi ini mengadopsi dan menggabungkan dua algoritma *Bayes* untuk merumuskan algoritma yang lebih baik yang dikenal sebagai *Hybrid Naïve Bayes Classifier* untuk model klasifikasi penjual benih ikan air tawar. Algoritma yang tergabung adalah *Gaussian Naïve Bayes* dan *Bernoulli Naïve Bayes*.

2.6. Confusion Matrix

Confusion Matrix merupakan media yang digunakan untuk menganalisis berbagai model yang dapat mengenali tupel pada kelas yang berbeda [19]. Evaluasi dengan fungsi matriks menghasilkan nilai akurasi, presisi, dan *recall*. Matriks ini terdiri dari dua kelas yaitu kelas yang dianggap positif dan kelas yang dianggap negatif [20][21]. Pada tabel 4 merupakan gambaran *Confusion Matrix* [21].

Tabel 3. *Confusion Matrix*

Nilai Prediksi	Nilai Aktual		
		Positif	Negatif
	Positif	TP	FP
Negatif	FN	TN	

1. Nilai akurasi menunjukkan akurasi model dalam mengklasifikasikan data dengan baik [21]. Berikut ini adalah formula dari akurasi *Confusion Matrix* :

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

2. Nilai presisi menunjukkan ketepatan antara data yang diinginkan dengan hasil prediksi yang diberikan oleh model [21]. Berikut ini adalah formula dari presisi *Confusion Matrix* :

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

3. Nilai *recall* atau sensitivitas: menggambarkan keberhasilan model dalam mengambil sepotong informasi [21]. Berikut ini adalah formula dari *recall Confusion Matrix* :

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

3. HASIL DAN PEMBAHASAN

3.1. Clustering penjual benih ikan berdasarkan fitur-fitur yang ada.

Data yang sudah diambil dari *e-marketplace* ini dikumpulkan, dibersihkan, dan diimplementasikan dengan *Gaussian Mixtures Clustering* untuk mengidentifikasi dan mengkategorikan kecenderungan rekomendasi para penjual benih ikan ini. Tabel 4 di bawah ini menunjukkan sebaran dari penjual benih ikan dalam hal kecenderungan rekomendasi mereka.

Tabel 4. Penambahan Label

Bibit	Clarias	Clarias	...	Bramidae
uk2_3	150	150	...	450
uk3_4	150	150	...	750
uk4_6	220	200	...	1000
uk7_9	450	450	...	1200
uk9_12	600	500	...	3200
Remaja	1100	1100	...	7000
Pengiriman	Jauh	Jauh	...	Jauh
Diskon	1	1	...	1
Jaminan	Ya	Ya	...	Tidak
Jumlah Penilaian	Rendah	Sedang	...	Rendah
Label	Tidak	Ya	...	Tidak

Untuk lebih jauh memvalidasi apakah hasil *clustering* itu valid, peneliti melaporkan hasil ini ke pembudidaya ikan. Sebanyak 30 orang pembudidaya yang ada di Semarang dan sekitarnya setuju dengan hasil kategori atau kecenderungan rekomendasi penjual benih ikan.

3.2. Hibridisasi Algoritma Naïve Bayes

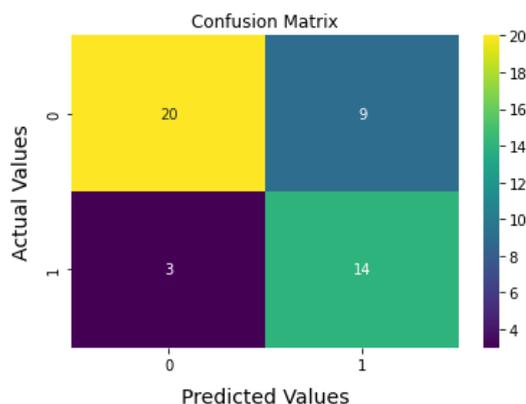
Pada proses hibridisasi ini, peneliti menggunakan algoritma *Gaussian Naïve Bayes* dan *Bernoulli Naïve Bayes*. Pada proses ini dihasilkan probabilitas dari kedua algoritma yang nantinya akan digunakan sebagai data *hybrid* dan bisa dilihat pada Tabel 5.

Tabel 5. Data Probabilitas

Gaussian NB	Bernoulli NB
(0.987219 , 0.012781)	(0.832841 , 0.167159)
(0.991663 , 0.008337)	(0.017784 , 0.982216)
(0.960761 , 0.039239)	(0.042767 , 0.957233)
(0.971451 , 0.028549)	(0.946165 , 0.053835)
(0.957791 , 0.042209)	(0.795180 , 0.204820)
.....
(0.185255 , 0.814745)	(0.892255 , 0.107745)
(0.236906 , 0.763094)	(0.892255 , 0.107745)
(0.185255 , 0.814745)	(0.865824 , 0.134176)
(0.184443 , 0.815557)	(0.865824 , 0.134176)
(0.320041 , 0.679959)	(0.892255 , 0.107745)

3.3. Performa Algoritma Hybrid Naïve Bayes

Gambar 2, 3, dan 4 menunjukkan hasil kinerja yang berbeda dari model yang berbeda dihasilkan dalam evaluasi klasifikasi. Ada 3 model yang berbeda dibandingkan : Model A, hanya menggunakan algoritma *Gaussian Naïve Bayes* dan telah diimplementasikan dengan 10 kali lipat validasi silang; Model B, hanya menggunakan algoritma *Bernoulli Naïve Bayes* dan telah diimplementasikan dengan 10 kali lipat validasi silang; Model C, menggunakan algoritma *Hybrid Naïve Bayes* dan telah diimplementasikan dengan 10 kali lipat validasi silang. Semua model diimplementasikan di *Jupyter Notebook*.

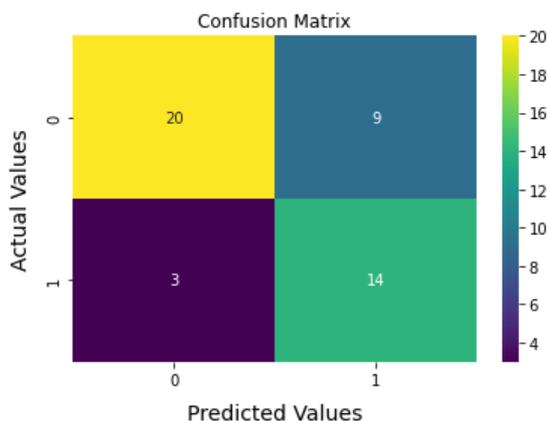


Gambar 4. Hybrid NB Confusion Matrix

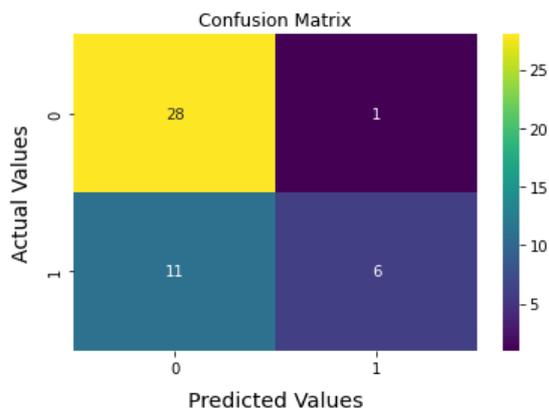
Untuk tujuan perbandingan, dari Gambar 2, 3, dan 5, peneliti memasukkan Model A, B, dan C ke dalam Tabel 6 yang sudah dianalisis dari akurasi data *training* dan data *testing* sehingga didapatkan model yang menjadi model terbaik.

Tabel 6. Perbandingan Model

	Testing	Training
Model A	73,91%	78,26%
Model B	73,91%	65,21%
Model C	73,91%	82,61%



Gambar 2. Gaussian NB Confusion Matrix



Gambar 3. Bernoulli NB Confusion Matrix

Mempertimbangkan Model C sebagai model terbaik, analisis *Confusion Matrix* ditunjukkan pada Tabel 6. Dari 46 prediksi pada pembuatan model menghasilkan 34 dari prediksi tersebut adalah prediksi yang benar. Di antara berbagai kelas atau kecenderungan rekomendasi, dapat dilihat pada *Confusion Matrix* yang hanya sedikit kasus salah diklasifikasikan oleh model. Jika model akan digunakan secara khusus untuk mengidentifikasi modelnya dapat memberikan hasil yang cukup baik sebagai kesalahan klasifikasi yang lebih rendah diamati di 2 kelas tersebut. Sedangkan pada Tabel 7, peneliti menjabarkan tentang evaluasi yang dilakukan dengan *Confusion Matrix* pada Model C. Berikut ini adalah proses perhitungan untuk menghitung akurasi, presisi dan *recall*-nya.

Tabel 7. Evaluasi Model

	Akurasi	Presisi	Recall	Training
Model C	73,91%	60,86%	82,35%	82,61%

Setelah itu, untuk memastikan nilai prediksi untuk rekomendasi atau tidaknya penjual benih ikan, peneliti melakukan uji untuk nilai prediksi positif dan negatif. Uji ini dilakukan agar prediksi yang dilakukan dapat berjalan baik dan sesuai dengan fungsinya. Berikut ini adalah perhitungan untuk uji prediksi :

1. Nilai uji prediksi positif :

$$NPP = \frac{(TP)}{(TP+FP)} = \frac{14}{(14+9)} = 0,6086 \quad (4)$$

Nilai prediktif positif menunjukkan probabilitas bahwa penjual benih ikan akan memiliki ciri-ciri sesuai dengan fitur jika hasil prediksinya rekomendasi. NPP sebesar 60,86% berarti jika hasil prediksi rekomendasi maka probabilitas penjual benih ikan yang benar-benar rekomendasi adalah 60,86%.

2. Nilai uji prediksi negatif:

$$NPN = \frac{(TN)}{(TN+FN)} = \frac{20}{(20+3)} = 0,8695 \quad (5)$$

Nilai prediktif negatif menunjukkan probabilitas bahwa penjual benih ikan akan memiliki ciri-ciri sesuai jika hasil prediksinya tidak rekomendasi. NPN sebesar 86,95% berarti jika hasil prediksi tidak rekomendasi maka probabilitas penjual benih ikan yang benar-benar tidak rekomendasi adalah 86,95%.

Secara keseluruhan, model lebih besar dalam mengidentifikasi penjual benih ikan yang diharapkan lebih banyak tidak rekomendasi, atau dalam mengidentifikasi penjual benih ikan yang tidak direkomendasikan. Jika seorang pembudidaya ikan air tawar menggunakan model ini sebagai alat untuk deteksi dini dalam pencegahan penjudul benih ikan yang tidak direkomendasikan, maka model ini pasti dapat memenuhi tujuannya.

4. DISKUSI

Salah satu penelitian yang pernah dilakukan adalah implementasi metode *Hybrid Naïve Bayes Classifier* melakukan klasifikasi model sentimen analisis pada *Twitter* tentang pasar saham untuk membandingkan akurasi, presisi, *recall*, dan *f-1 score* dengan metode *Multinomial Naïve Bayes Classifier* [15]. Hasilnya metode *Hybrid Naïve Bayes Classifier* dalam akurasi, presisi, *recall*, dan *f-1 score* dengan perbandingan 90% : 83% (*Hybrid Naïve Bayes* : *Multinomial Naïve Bayes*). Dalam penelitian yang lain dengan metode *Gaussian-Bernoulli Mixed Naïve Bayes (Hybrid Naïve Bayes)* untuk memprediksi kecenderungan prokrastinasi akademik siswa dalam pembelajaran matematika daring. Penelitian ini juga membuktikan bahwa seleksi fitur dan hibridisasi algoritma *Naïve Bayes* dapat meningkatkan akurasi dari 64% ke 91% [22]. Dari kedua penelitian diatas menunjukkan bahwa algoritma memiliki hasil klasifikasi yang cukup baik. Oleh karena itu, peneliti ingin menguji coba pada model klasifikasi untuk pemodelan rekomendasi penjual benih ikan dengan membandingkan algoritma *Hybrid Naïve Bayes* dengan algoritma *Gaussian Naïve Bayes* dan *Bernoulli Naïve Bayes*. Analisis yang dilakukan oleh peneliti juga sudah divalidasi oleh pembudidaya sehingga dataset yang digunakan merupakan dataset yang valid.

5. KESIMPULAN

Berdasarkan temuan penelitian ini, peneliti menyimpulkan bahwa : *Hybrid Naïve Bayes* yang dihasilkan model *Naïve Bayes* dapat dimanfaatkan oleh pembudidaya ikan air tawar untuk memprediksi rekomendasi penjual benih ikan. Proses *training* yang dilakukan oleh algoritma *Hybrid Naïve Bayes* terbukti jauh lebih baik dalam memodelkan klasifikasi untuk rekomendasi penjual benih ikan air tawar daripada algoritma *Gaussian Naïve Bayes* dan *Bernoulli Naïve Bayes*.

Hasil prediksi dari alat ini dapat digunakan oleh pembudidaya ikan air tawar dalam melakukan pengambilan keputusan dan perencanaan jangka panjang. Peneliti merekomendasikan model ini karena *Naïve Bayes* hanyalah satu diantara banyak model Machine Learning yang tersedia, menyelidiki algoritma yang lebih baru dan canggih, seperti *Neural Networks* dan lainnya, dengan lebih banyak data atau masukan fitur yang mungkin juga memiliki nilai prediktif yang bagus untuk merekomendasikan penjual benih ikan.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, "Pencarian Produksi Ikan Dengan Perbandingan Tahun", 2022. <https://statistik.kkp.go.id/home.php?m=total&i=2#panel-footer> [accessed Oct. 2, 2022].
- [2] L. Lukmandono, M. Basuki, M. J. Hidayat, V. Setyawan, "Pemilihan Supplier Industri Manufaktur Dengan Pendekatan AHP dan TOPSIS," *OPSI*, vol. 12, no. 2, pp. 83-88, 2019, DOI : <https://doi.org/10.31315/opsi.v12i2.3146>.
- [3] V. B. Schramm, L. P. B. Cabral, F. Schramm, "Approaches for supporting sustainable supplier selection-A literature review," *Journal of cleaner production*, vol. 273, pp. 123089, 2020, DOI : <https://doi.org/10.1016/j.jclepro.2020.123089>
- [4] A. Balango, S. Risnanto, P. Mauliana, W. Wigun, "Aplikasi Seleksi Pemasok Konstruksi Menggunakan Metode Weighted Product Berbasis Mobile Di PT Binarthama Kharisma," *Infotronik: Jurnal Teknologi Informasi dan Elektronika*, vol. 5, no. 1, pp. 1-10, 2020, DOI : <https://doi.org/10.32897/infotronik.2020.5.1.34>, DOI : <https://doi.org/10.32897/infotronik.2020.5.1.354>
- [5] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, 2020, DOI : <https://doi.org/10.1016/j.aci.2018.08.003>.

- [6] A. H. Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris," *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, vol. 7, no. 2, pp. 45-51, 2021, DOI : <https://doi.org/10.35329/jiik.v7i2.203>
- [7] E. A. Sambudi, "Analisa Pemilihan Supplier dengan Metode Analytic Hierarchy Process: Kasus Perusahaan Otomotif di Sunter," *Jurnal Operations Excellence: Journal of Applied Industrial Engineering*, vol. 11, no. 3, pp. 322-329, 2019, DOI : [10.22441/oe.v11i3.2019.040](https://doi.org/10.22441/oe.v11i3.2019.040)
- [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] J. W. Gullifer, and D. Titone, "Engaging proactive control: Influences of diverse language experiences using insights from machine learning", *Journal of Experimental Psychology: General*, vol.150, no.3, pp. 414, 2021, DOI : <https://doi.org/10.1037/xge0000933>.
- [10] S. García, J. Luengo, F. Herrera, "Data preprocessing in data mining," *Cham, Switzerland: Springer International Publishing*, vol. 72, pp. 59-139, 2015.
- [11] R. A. R. Mahmood, A. Abdi, M. Hussin, "Performance Evaluation of Intrusion Detection System using Selected Features and Machine Learning Classifiers," *Majallat Baghdād Lil-'ulūm*, vol. 18, no.2, pp. 884, 2021, DOI : [https://doi.org/10.21123/bsj.2021.18.2\(Suppl.\).0884](https://doi.org/10.21123/bsj.2021.18.2(Suppl.).0884)
- [12] Z. M. Mosa, N. H. Ghaeb, A. H. Ali, "Detecting keratoconus by using SVM and decision tree classifiers with the aid of image processing," *Baghdad Science Journal*, vol. 16, no. 4, 2019, DOI : [https://doi.org/10.21123/bsj.2019.16.4\(Suppl.\).1022](https://doi.org/10.21123/bsj.2019.16.4(Suppl.).1022)
- [13] W. Jia, Y. Tan, L. Liu, J. Li, H. Zhang, K. Zhao, "Hierarchical prediction based on two-level Gaussian mixture model clustering for bike-sharing system," *Knowledge-Based Systems*, vol. 178, pp. 84-97, 2019, DOI : <https://doi.org/10.1016/j.knosys.2019.04.020>
- [14] E. Patel, D. S. Kushwaha, "Clustering cloud workloads: K-means vs gaussian mixture model," *Procedia computer science*, vol. 171, pp. 158-167, 2020, DOI : <https://doi.org/10.1016/j.procs.2020.04.017>
- [15] G. A. A. J. Alkubaisi, S. S. Kamaruddin, H. Husni. "Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers," *Comput. Inf. Sci.*, vol.11, no.1, pp.52-64, 2018.
- [16] M. Rashid, A. Hamid, S. A. Parah, "Analysis of streaming data using big data and hybrid machine learning approach," *In Handbook of Multimedia Information Security: Techniques and Applications*, pp. 629-643, 2019, DOI : [10.1007/978-3-030-15887-3_30](https://doi.org/10.1007/978-3-030-15887-3_30)
- [17] J. Li, Z. Li, C. Wen, Q. Peng, P. Huang, "Train operation conflict detection for high-speed railways: a Naïve Bayes approach," *International Journal of Rail Transportation*, pp. 1-19, 2022, DOI : <https://doi.org/10.1080/23248378.2022.2071346>
- [18] V. R. Balaji, S. T. Suganthi, R. Rajadevi, V. K. Kumar, B. S. Balaji, S. Pandiyan, "Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naive Bayes Classifier," *Measurement*, vol. 163, pp. 107922, 2020, DOI : <https://doi.org/10.1016/j.measurement.2020.107922>
- [19] E. P. Cynthia, E. Ismanto, "Metode Decision Tree Algoritma C. 45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji," *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, vol. 3, pp. 1-13, 2018.
- [20] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *Jurnal Teknoinfo*, vol. 14, no. 2, pp. 115-123, 2020, DOI: <https://doi.org/10.33365/jti.v14i2.679>
- [21] J. Xu, Y. Zhang, D. Miao, "Three-way confusion matrix for classification: A measure driven view" *Information sciences*, vol. 507, pp. 772-794, 2020, DOI : <https://doi.org/10.1016/j.ins.2019.06.064>
- [22] Charles D. O. Godinez, L. S. Lomibao, "A Gaussian-Bernoulli Mixed Naïve Bayes Approach to Predict Students' Academic Procrastination Tendencies in Online Mathematics Learning," *American Journal of Educational Research*, vol. 10, no. 4, pp. 223-232, 2022, DOI : [10.12691/education-10-4-10](https://doi.org/10.12691/education-10-4-10)