

Penerapan Algoritma Naïve Bayes Berbasis *Backward Elimination* Untuk Prediksi Pemesanan Kamar Hotel

Haditsah Annur

Teknik Informatika, Universitas Ichsan Gorontalo, Gorontalo, Indonesia
haditsah@unisan.ac.id

Abstrak— Perhotelan merupakan salah satu badan usaha yang memiliki potensi untuk tumbuh dan berkembang yang membutuhkan banyak investasi yang berdiri baik di tengah kota maupun di daerah tujuan wisata. Pemesanan kamar hotel dapat dilakukan oleh pelanggan sebelum mereka menggunakan hotel melalui media online, namun permasalahan yang terjadi adalah pelanggan yang telah melakukan pemesanan kamar dapat membatalkan pesanan dengan berbagai alasan, sehingga pelaku bisnis perhotelan merasa rugi karena pembatalan tersebut dapat memberikan peluang bagi desainer untuk mendapatkan pelanggan ini. Penelitian ini menggunakan Algoritma Naïve Bayes sebagai algoritma yang dapat menghasilkan akurasi tinggi dan dapat mengolah banyak data, serta pemilihan mundur sebagai pemilihan nilai parameter yang sesuai untuk meningkatkan akurasi. Penelitian ini menggunakan 10.000 data pelanggan kamar hotel dengan hasil akurasi menggunakan Algoritma Naïve Bayes 89.67%, dan hasil akurasi menggunakan Algoritma Naïve Bayes dan Pemilihan Fitur Backward Elimination sebesar 97.83%. Hasil Prediksi Check-Out, Dibatalkan dan No-Show.

Kata kunci: Prediksi, Pemesanan Kamar Hotel, Naïve Bayes, Backward Elimination.

Abstract— Hotel is one of the business entities that have the potential to grow and develop which requires a lot of investment that stands both in the middle of the city and in tourist destination areas. Hotel room reservations can be made by customers before they use the hotel via online media, but the problem that occurs is that customers who have made room bookings can cancel orders for various reasons, so that hoteliers feel a loss because the cancellation can provide opportunities for designers to get these customers. This study uses the Naïve Bayes Algorithm as an algorithm that can produce high accuracy and can process a lot of data, and backward selection as the selection of suitable parameter values to improve accuracy. This study uses 10,000 hotel room customer data with accuracy results using the Naïve Bayes Algorithm of 89.67%, and accuracy results using the Naïve Bayes Algorithm and Backward Elimination Feature Selection of 97.83%. Prediction Results Check-Out, Canceled and No-Show.

Keywords: Prediction, Booking Room Hotel, Naïve Bayes, Backward Elimination.

1. PENDAHULUAN

Hotel merupakan salah satu badan usaha yang mempunyai potensi untuk dapat tumbuh dan berkembang yang membutuhkan investasi yang tidak sedikit yang berdiri baik ditengah kota maupun di daerah-daerah tujuan wisata [1]. Usaha di bidang perhotelan semakin ketan persaingannya dikarenakan beberapa hal seperti kualitas, harga kamar dan pelayanan dari hotel tersebut [2]. Pelanggan berperan sangat penting dalam usaha di bidang perhotelan, seperti pelanggan hotel melakukan pemesanan kamar sebelum mereka menggunakan kamar hotel, salah satunya dengan cara menggunakan pemesanan media *online* sehingga pelanggan dapat terbantu dalam pemesanan kamar hotel. Dengan banyaknya pilihan hotel, maka pelanggan semakin selektif dalam memperhitungkan kualitas layanan yang ditawarkan oleh tiap-tiap hotel. Masalah yang terjadi adanya pelanggan yang telah melakukan pemesanan kamar dapat membatalkan pesanan dengan berbagai alasan, sehingga para pelaku usaha di bidang perhotelan merasakan kerugian dikarenakan pembatalan tersebut dapat memberikan kesempatan pesain untuk mendapatkan pelanggan tersebut.

Penelitian ini akan melakukan implementasi Algoritma *Naïve bayes* dengan seleksi fitur *Backward Elimination* dengan menggunakan tools Rapid Miner untuk mengetahui tingkat akurasi algoritma tersebut terhadap data pelanggan hotel, data yang digunakan di penelitian ini menggunakan data public www.kaggle.com dari penelitian “Hotel booking demand datasets”[3].

Algoritma *Naïve Bayes* merupakan pengklasifikasian statistik sebagai algoritma yang dapat menghasilkan akurasi yang tinggi dan dapat mengolah data yang banyak dan *Naïve bayes* merupakan algoritma algoritma yang mudah diimplementasikan sehingga sangat efektif apabila di uji dengan menggunakan dataset yang tepat, terutama bila *naïve bayes* dengan seleksi fitur, maka *naïve bayes* dapat mengurangi *redundant* pada data [4]. Selain itu juga Algoritma tersebut dapat dikategorikan sebagai Algoritma yang sederhana dan memiliki atribut bersifat berdiri sendiri sehingga dapat mengungguli algoritma yang lain [5].

Backward Elimination menghasilkan performa kinerja lebih baik Ketika dibandingkan dengan cara statistic dalam menyeleksi fitur. Kinerja terbaik dapat diperoleh dari sensitivitas, spesifitas dan keakuratan [6].

Selain itu juga dengan seleksi fitur tingkat akurasi yang dihasilkan meningkat, dengan memilih atribut yang relevan dan menghasilkan model terbaik [7]. *Backward Elimination* juga memiliki fungsi mengoptimalkan kinerja suatu model dengan sistem pemilihan dengan cara memilih variabel kedepan yaitu menguji semua variabel kemudian menghapus variabel yang dianggap tidak relevan [8].

2. METODE

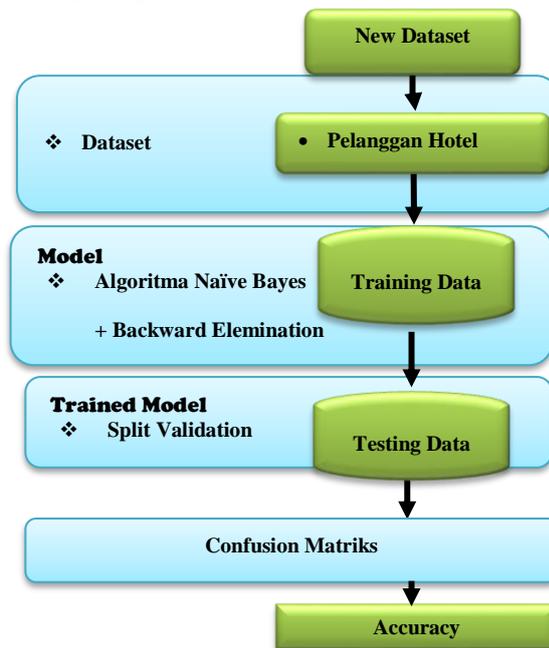
Penelitian ini menggunakan jenis penelitian eksperimen, pengumpulan data merupakan Langkah awal pada suatu penelitian. Data yang digunakan adalah dataset data publik www.kaggle.com. Jumlah dataset yang digunakan sebanyak 10.000 record tentang data pemesanan kamar hotel. Contoh dataset terdapat pada gambar di bawah ini :

Row No.	reservation...	hotel	is_canceled	arrival_date...	arrival_date...	arrival_date...	arrival_date...	stays_in_...
1	Check-Out	Resort Hotel	0	2015	July	27	1	0
2	Check-Out	Resort Hotel	0	2015	July	27	1	0
3	Check-Out	Resort Hotel	0	2015	July	27	1	0
4	Check-Out	Resort Hotel	0	2015	July	27	1	0
5	Check-Out	Resort Hotel	0	2015	July	27	1	0
6	Check-Out	Resort Hotel	0	2015	July	27	1	0
7	Check-Out	Resort Hotel	0	2015	July	27	1	0
8	Check-Out	Resort Hotel	0	2015	July	27	1	0
9	Canceled	Resort Hotel	1	2015	July	27	1	0
10	Canceled	Resort Hotel	1	2015	July	27	1	0

Gambar 1. Contoh Dataset

2.1 Desain Eksperimen

Model yang diusulkan dapat dilihat pada gambar di bawah ini :



Gambar 2. Desain Eksperimen

2.2 Algoritma Naïve Bayes

Algoritma *Naïve Bayes* digunakan untuk prediksi peluang atau kemungkinan suatu kelas, mempunyai pendapat yang kuat untuk tidak terkait pada masing-masing kondisi [9]. Selain itu juga Algoritma *Naïve Bayes* merupakan Teknik prediksi berdasarkan kemungkinan sederhana yang banyak digunakan disebabkan kecepatan dan kesederhanaannya [10].

Persamaan dari Algoritma *Naïve Bayes* sebagai berikut [11][12]:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Keterangan :

- X = Sampel data yang memiliki kelas (label yang tidak diketahui)
- H = Hipotesa bahwa X adalah data kelas (label)
- P(H) = Peluang dari hipotesa H
- P(X) = Peluang dari data sampel yang diamati
- P(X|H) = Peluang dari data sampel X bila diasumsikan bahwa hipotesa benar

2.3 Seleksi Fitur Backward Elimination

Seleksi fitur merupakan kegiatan yang dapat dilakukan secara *preprocessing* dan tujuannya memilih fitur yang berpengaruh dan mengesampingkan fitur yang tidak berpengaruh dalam pemodelan atau analisa data. Backward Elimination salah satu seleksi fitur yang digunakan sebagai metode pemilihan berfungsi mengeluarkan variabel satu per satu yang tidak signifikan atau tidak relevan [13][14]. Selain itu juga *Backward Elimination* adalah metode untuk mengidentifikasi fitur dalam kumpulan ata yang sama pentingnya dan membuang semua fitur lain yang tidak relevan dan berlebihan [15].

2.4 Confusion Matrix

Confusion Matrix merupakan Teknik pengujian yang menghasilkan evaluasi model dengna menggunakan tabel matriks, yang terdiri dari dua kelas yaitu kelas yang dianggap sebagai positif dan kelas yang dianggap sebagai negatif [16]. Selain itu juga untuk mengukur kinerja dapat menggunakan *Confusion Matrix Multiclass* seperti pada tabel di bawah ini [17] :

Tabel 1. Model *Confusion Matrix Multiclass*

		Prediksi		
		Class 1	Class 2	Class 3
Aktual	Class 1	TP	E ₁₂	E ₁₃
	Class 2	E ₂₁	TP	E ₂₃
	Class 3	E ₃₁	E ₃₂	TP

Persamaan tingkat akurasi pada matriks digunakan pada rumus [13] :

$$accuracy = \frac{TP}{jumlah\ data}$$

Catatan bahwa matriks *confusion multiclass* adalah perkembangan dari matriks *confusion binary* dimana sebelumnya terdapat FN (*False Negative*), FP (*False Positive*), dan TN (*True Negative*). Kemudian pada matriks *confusion multiclass* hanya tertera TP (*True Positive*) karena untuk penentuan FN adalah dari seluruh jumlah baris per variabel sedangkan untuk penentuan FP adalah dari seluruh julah kolom per variable dan TN adalah kasus-kasus dimana kita prediksi tidak ada dan nilai aktualnya salah [18].

3. HASIL DAN PEMBAHASAN

3.1 Pengujian menggunakan Algoritma Naïve Bayes

Pengujian menggunakan Algoritma *Naïve Bayes*. Hasil pengujian akurasi *Confusion Matrix Multiclass* dapat di lihat pada gambar di bawah ini :

accuracy: 89.67%

	true Check-Out	true Canceled	true No-Show	class precision
pred. Check-Out	3666	0	0	100.00%
pred. Canceled	0	5143	2	99.96%
pred. No-Show	0	1031	158	13.29%
class recall	100.00%	83.30%	98.75%	

Gambar 3.1 Hasil Akurasi Algoritma *Naïve Bayes*

Dari gambar di atas diperoleh tingkat nilai akurasi menggunakan Algoritma *Naïve Bayes* sebesar 89,67%, dan dapat dihitung pada persamaan di bawah ini :

$$Accuracy = \frac{3666 + 5143 + 158}{10000} = \frac{8967}{10000} = 0,8967$$

3.2 Pengujian menggunakan Algoritma *Naïve Bayes* Berbasis *Backward Elimination*

Pengujian dilakukan model 10 *folds split validation* menggunakan seleksi fitur *Backward Elimination*., Hasil pengujian akurasi *Confusion Matrix Multiclass* dapat dilihat pada gambar di bawah ini :

accuracy: 98.50%				
	true Check-Out	true Canceled	true No-Show	class precision
pred. Check-Out	353	0	0	100.00%
pred. Canceled	0	628	6	99.05%
pred. No-Show	0	9	4	30.77%
class recall	100.00%	98.59%	40.00%	

Gambar 3.2 Hasil akurasi Algoritma *Naïve Bayes* dengan Seleksi Fitur *Backward Elimination*

Dari gambar di atas diperoleh tingkat akurasi menggunakan Algoritma *Naïve Bayes* dengan seleksi fitur *Backward Elimination* sebesar 98,50 % dan dapat dihitung pada persamaan di bawah ini :

$$Accuracy = \frac{353 + 628 + 4}{1000} = \frac{985}{1000} = 0,985$$

3.3 Pembahasan

Hasil akurasi dari model yang dihasilkan dengan Algoritma *Naïve Bayes* dan *Naïve Bayes* berbasis *Backward Elimination* terlihat pada perbandingan nilai *accuracy* yang dirangkumkan dalam tabel di bawah ini :

Tabel 2.

Hasil Akurasi Algoritma *Naïve Bayes* dan Algoritma *Naïve Bayes* Berbasis *Backward Elimination*

	<i>Naïve Bayes</i>	<i>Naïve Bayes + Backward Elimination</i>
<i>Accuracy</i>	89,67%	98,50%

Tabel 2. Menampilkan nilai akurasi Algoritma *Naïve Bayes* berbasis *Backward Elimination* sebesar 98,50 %, dengan demikian bahwa nilai klasifikasi sangat baik, dari hasil tersebut diketahui dalam memprediksi dapat digunakan, karena proses seleksi fitur dari *Backward Elimination* menghasilkan atribut-atribut yang berpengaruh.

Waktu yang digunakan untuk memproses pengujian membutuhkan waktu kurang dari 1 menit dengan 10000 record yang digunakan, hal ini menunjukkan semakin besar jumlah record yang digunakan, maka waktu pengujian mempengaruhi hasil akurasi yang didapatkan.

4. KESIMPULAN

Kesimpulan penelitian ini hasil pengujian dari Algoritma *Naïve Bayes* berbasis *Backward Elimination* diperoleh nilai *accuracy* 98,50%. maka algoritma tersebut merupakan salah satu algoritma yang baik digunakan untuk memprediksi pemesanan kamar hotel, oleh karena itu usaha perhotelan dapat mencegah pelanggan tidak berpindah ke hotel yang lain.

5. DAFTAR PUSTAKA

[1] E. Winata, S. Tinggi, and I. M. Sukma, “Pengaruh Kepuasan Kerja dan Kompensasi terhadap Kinerja Karyawan pada Hotel Inna Dharma Deli Medan,” *J. Ilman*, vol. 4, no. 1, pp. 1–17, 2016.
[2] E. E. Singon, I. Elim, D. Afandi, F. Ekonomi, J. Akuntansi, and U. S. Ratulangi, “ANALISIS PENENTUAN HARGA JUAL KAMAR HOTEL PADA HOTEL GRAND PURI MANADO

- ANALYSIS OF DETERMINING THE SELLING PRICE OF HOTEL ROOMS IN GRAND PURI MANADO HOTEL,” vol. 7, no. 3.
- [3] N. Antonio, A. de Almeida, and L. Nunes, “Hotel booking demand datasets,” *Data Br.*, vol. 22, pp. 41–49, 2019.
- [4] R. Satria, “Integrasi Bagging dan Greedy Forward Selection pada Prediksi Cacat Software dengan Menggunakan Naive Bayes,” *Integr. Bagging dan Greedy Forw. Sel. pada Prediksi Cacat Softw. dengan Menggunakan Naive Bayes*, vol. 1, no. 2, pp. 101–108, 2015.
- [5] M. Hasan, “Menggunakan Algoritma Naive Bayes Berbasis,” vol. 9, pp. 317–324, 2017.
- [6] A. Bode, “K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika,” *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 188–195, 2017.
- [7] I. Herliawan *et al.*, “Classification of Liver Disease By Applying Random Forest Algorith and Backward Elimination,” *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komput.*, vol. 6, no. 1, pp. 89–94, 2020.
- [8] W. Nugroho, “Optimasi Metode K-Nearest Neighbours dengan Backward Elimination Menggunakan Dataset Software Effort Estimation Bianglala Informatika,” vol. 8, no. 2, pp. 129–133, 2020.
- [9] . N., N. Septianti, N. Retnowati, and A. Wibowo, “Prediksi Tingkat Kelulusan Tepat Waktu Mahasiswa Menggunakan Algoritma Naive Bayes pada Universitas XYZ,” *Ultim. J. Tek. Inform.*, vol. 12, no. 2, pp. 104–107, 2020.
- [10] M. Guntur, J. Santony, and Y. Yuhandri, “Prediksi Harga Emas dengan Menggunakan Metode Naive Bayes dalam Investasi untuk Meminimalisasi Resiko,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 1, pp. 354–360, 2018.
- [11] I. Muhammad, “Implementasi Data Mining Dengan Algoritma Naive Bayes Untuk Memprediksi Angka Kelahiran,” *J. Pelita Inform.*, vol. 7, no. 3, pp. 421–428, 2019.
- [12] H. Annur, “Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes,” *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018.
- [13] A. D. Ghani, N. Salman, and Mustikasari, “Algoritma k-Nearest Neighbor Berbasis Backward Elimination Pada Client Telemarketing,” *Pros. Semin. Ilm. Sist. Inf. dan Teknol. Inf.*, vol. 8, no. 2, pp. 141–150, 2019.
- [14] Achmad Saiful Rizal and Moch. Lutfi, “Prediksi Hasil Pemilu Legislatif Menggunakan Algoritma K-Nearest Neighbor Berbasis Backward Elimination,” *J. Resist. (Rekayasa Sist. Komputer)*, vol. 3, no. 1, pp. 27–41, 2020.
- [15] F. Ma’arif and T. Arifin, “Optimasi Fitur Menggunakan Backward Elimination Dan Algoritma SVM Untuk Klasifikasi Kanker Payudara,” *J. Inform.*, vol. 4, no. 1, pp. 46–53, 2017.
- [16] D. Salmu, S. and A. Solichin, “Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta Prediction of Timeliness Graduation of Students Using Naive Bayes : A Case Study at Islamic State University Syarif Hidayatullah Jakarta,” *Pros. Semin. Nas. Multidisiplin Ilmu*, no. April, pp. 701–709, 2017.
- [17] D. A. Fauziah *et al.*, “Klasifikasi Berita Politik Menggunakan Algoritma K-nearst Neighbor (Classification of Political News Content using K-Nearest Neighbor) Abstrak,” *J. Sist. Inf. Univ. Jember*, vol. 6, no. 2, p. 8, 2018.
- [18] V. W. Siburian and I. E. Mulyana, “Prediksi Harga Ponsel Menggunakan Metode Random Forest,” *Pros. Annu. Res. Semin.*, vol. 4, no. 1, pp. 144–147, 2018.