

ANALISIS PENYEBAB TINGGINYA CANCELLATION RATE PADA RESORT HOTEL MENGGUNAKAN ALGORITMA C4.5

Fisca Shafa Salsabila¹, Apriade Voutama²

Program Studi Sistem Informasi, Fakultas Ilmu Komputer

Universitas Singaperbangsa Karawang

Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat
salsabilashafafisca@gmail.com¹, apriade.voutama@staff.unsika.ac.id²

Abstrak

Tingginya tingkat pembatalan reservasi hotel menjadi tantangan utama dalam industri perhotelan karena dapat berdampak pada pendapatan dan efisiensi operasional. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi pembatalan reservasi pada resort hotel dengan menggunakan algoritma C4.5. Data yang digunakan berasal dari platform Kaggle, terdiri atas 119.390 observasi dari dua jenis hotel dalam periode 2015–2017. Metode *Knowledge Discovery in Database* (KDD) diterapkan melalui tahapan seleksi, praproses, transformasi, penambangan data, dan evaluasi model. Hasil analisis menunjukkan bahwa *lead time* yang panjang, jenis deposit "No Deposit", dan tipe pelanggan individu merupakan faktor utama yang memengaruhi pembatalan. Kontribusi utama dari penelitian ini adalah penyediaan model prediktif yang akurat (akurasi 96,08%) yang dapat digunakan sebagai dasar pengambilan keputusan strategis dalam manajemen reservasi hotel.

Kata Kunci : Pembatalan Reservasi, Algoritma C4.5, Hotel, Knowledge Discovery in Database, Data Mining.

Abstract

The high rate of hotel reservation cancellations presents a major challenge in the hospitality industry, as it can impact both revenue and operational efficiency. This study aims to analyze the factors influencing reservation cancellations in resort hotels using the C4.5 algorithm. The data used in this research was obtained from the Kaggle platform and consists of 119,390 observations from two types of hotels over the period 2015–2017. The Knowledge Discovery in Database (KDD) methodology is applied through the stages of selection, preprocessing, transformation, data mining, and model evaluation. The analysis results indicate that long lead time, "No Deposit" type, and individual customer type are the main factors influencing cancellations. The primary contribution of this research is the provision of an accurate predictive model (96.08% accuracy), which can be used as a strategic decision-making tool in hotel reservation management.

Keyword : Reservation Cancellation, C4.5 Algorithm, Hotel, Knowledge Discovery in Database, Data Mining.

PENDAHULUAN

Adanya peningkatan jumlah wisatawan untuk keperluan bisnis ataupun liburan, membuat industri perhotelan mengalami pertumbuhan di dunia. Berbagai hotel terus meningkatkan kualitas layanan mereka karena adanya peningkatan permintaan layanan perhotelan yang semakin tinggi (Syafriamansyah, 2024). Selain menjadi penopang penting bagi sektor pariwisata, industri perhotelan juga memberikan kontribusi signifikan terhadap pertumbuhan ekonomi, pengembangan infrastruktur serta penciptaan lapangan kerja (Wisnawa, 2024). Pemikiran dan implemementasi dari perencanaan yang strategis sangat diperlukan untuk berada diposisi yang unggul dalam memenangkan persaingan pada industri perhotelan (Teguh & Ciawati, 2020).

Tingkat pembatalan reservasi yang tinggi menjadi tantangan besar dalam industri perhotelan terutama dalam pengoptimalan pendapatan serta efisiensi operasional. Berdasarkan data dari Asosiasi Hotel Indonesia (PHRI) pada tahun 2023, rata-rata pembatalan reservasi hotel di Indonesia mencapai 18%, dengan daerah wisata seperti Bali dan Yogyakarta mencatat tingkat pembatalan hingga 25%. Hal ini menunjukkan pentingnya pemahaman mendalam terhadap penyebab pembatalan untuk meminimalkan kerugian finansial, ketidakpastian pengelolaan kamar, serta meningkatkan kepuasan pelanggan. Dengan menganalisis pola pembatalan, penelitian ini

memiliki tujuan untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap pembatalan reservasi dan memberikan solusi berbasis data untuk membantu hotel dalam mengurangi tingkat pembatalan. Melalui proses analisis data menggunakan metode C4.5, strategi yang lebih efektif dapat dirancang untuk meningkatkan retensi pelanggan, mengoptimalkan alokasi kamar, serta meningkatkan efisiensi dalam sistem reservasi hotel.

Kegiatan berpikir yang menghasilkan komponen dari proses penguraian suatu keseluruhan sehingga dapat mengenal hubungan dan fungsi satu sama lain dalam keseluruhan yang terpadu yaitu arti dari analisis (Septiani et al., 2020). Kesimpulan dapat terpengaruh secara fatal akibat kesalahan dalam spesifikasi penganalisis, dan bahkan penggunaan serta penerapan hasil penelitian dapat berdampak lebih buruk (Millah et al., 2023). Algoritma C4.5 termasuk pada kelompok algoritma *decision tree*. *Input* dari algoritma ini yaitu berupa *training samples* dan *samples* (Purwanto et al., 2020). Dalam penanganan suatu permasalahan yang lebih dalam, algoritma C4.5 memiliki beberapa perubahan serta peningkatan yang membuat lebih efektif (Nazifah & Prianto, 2023). Pohon keputusan yang paling populer salah satunya yaitu algoritma C4.5. Kegunaan dari pohon keputusan yaitu untuk eksplorasi data serta mencari hubungan tersembunyi antara beberapa *variable target* dan *variable input* kandidat (Girsang et al., 2022).

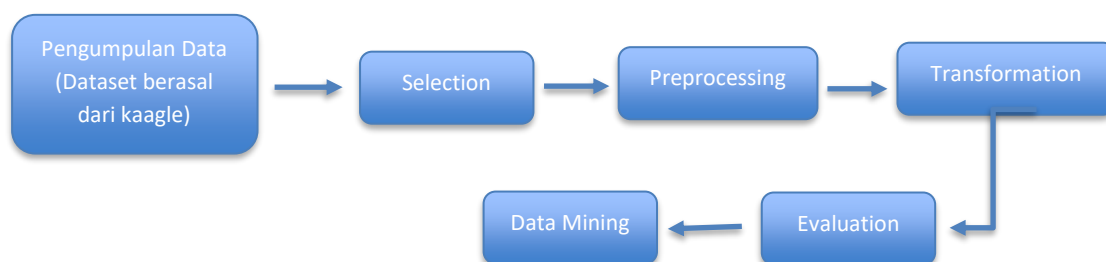
PENELITIAN RELEVAN

Penelitian sebelumnya sudah membahas beberapa faktor yang menyebabkan pembatalan transaksi dan strategi mitigasi untuk mengatasi permasalahan tersebut. (Arrosyid et al., 2025) dalam studi mereka menganalisis tingginya tingkat pembatalan reservasi pada City Hotel menggunakan metode *Issue Tree*. Dengan pendekatan ini, mereka mengidentifikasi dua kategori utama, yaitu faktor fasilitas hotel seperti meal plan, tipe kamar, dan lahan parkir, serta faktor pelayanan, termasuk saluran distribusi pemesanan dan lead time. Penelitian tersebut menyarankan peningkatan fasilitas dan kualitas pelayanan untuk menurunkan cancellation rate.

Di sisi lain, Penelitian oleh Nor'Aini & Rakhmawati (2024) membahas pembatalan dalam konteks yang berbeda, yakni perpajakan, sehingga relevansi langsung terhadap konteks perhotelan cukup terbatas. Namun, pendekatan mereka dalam mengidentifikasi kesalahan administratif sebagai faktor pembatalan tetap memberikan inspirasi metodologis. Dengan demikian, kontribusi ilmiah utama dari penelitian ini adalah penerapan algoritma C4.5 dalam konteks prediksi pembatalan reservasi pada resort hotel dengan akurasi tinggi (96,08%), serta pengembangan atribut yang lebih komprehensif dan relevan secara praktis bagi industri perhotelan. Penelitian ini tidak hanya mendalami penyebab pembatalan, tetapi juga menyajikan model prediktif yang dapat digunakan langsung oleh manajemen hotel untuk mengurangi tingkat pembatalan secara strategis.

METODE PENELITIAN

Metodologi yang diterapkan dalam penelitian ini adalah *Knowledge Discovery in Database* (KDD). KDD merupakan proses komputasi yang digunakan untuk mengekstraksi informasi dengan memanfaatkan perhitungan matematis serta melakukan kalkulasi probabilistik berdasarkan probabilitas suatu tindakan di masa mendatang (Veronica Agustin & Voutama, 2023).



Gambar 1. Algoritma C4.5

1. *Selection*

Pada tahap seleksi, atribut akan dipilih dan kemudian diintegrasikan hingga terbentuk sebuah dataset. Proses pembangunan dataset dianggap penting karena berperan dalam pembelajaran data mining serta penemuan pola baru yang didasarkan pada dataset yang telah dibentuk (Veronica Agustin & Voutama, 2023). Seleksi dilakukan berdasarkan beberapa kriteria, yaitu relevansi terhadap variabel target (pembatalan), tingkat kelengkapan data (minim missing value), kemudahan interpretasi, serta tidak bersifat identitas pribadi seperti nama, email, dan nomor telepon. Selain itu, atribut yang memiliki korelasi tinggi satu sama lain juga dipertimbangkan untuk disederhanakan.

2. *Preprocessing*

Tahap kedua melibatkan proses data *preprocessing* yang dilakukan menggunakan bahasa pemrograman *Python*. Pada tahap ini, beberapa data yang tidak relevan dihapus, duplikasi data dieliminasi, kesalahan data diperbaiki, serta penanganan terhadap *missing value* dan aspek lainnya diterapkan (Nugraha et al., 2024; Veronica Agustin & Voutama, 2023).

3. *Transformation*

Pada tahap ini mencakup normalisasi atau standarisasi, di mana seluruh atribut distandarisasi agar setiap atribut memiliki kontribusi yang proporsional terhadap hasil akhir dalam proses data mining (Nugraha et al., 2024).

4. *Data Mining*

Langkah selanjutnya adalah tahap data mining, pencarian pola atau ekstraksi informasi dilakukan dari data yang telah dipilih menggunakan algoritma C4.5. Proses ini dilakukan dengan memanfaatkan bahasa pemrograman *Python* melalui *platform Google Colaboratory*. Dalam penelitian ini, algoritma C4.5 digunakan sebagai metode utama. (Veronica Agustin & Voutama, 2023).

5. *Evaluation*

Tahap ini dilakukan untuk menafsirkan serta menginterpretasikan pola yang dihasilkan dari proses data mining. Informasi yang sudah diperoleh harus disajikan dalam bentuk yang mudah dipahami. Dalam tahap ini, terjadi pemeriksaan terhadap pola serta informasi yang mungkin bertentangan dengan hipotesis maupun fakta yang sudah ada (Veronica Agustin & Voutama, 2023).

HASIL DAN PEMBAHASAN

Pada penelitian ini, analisis data dilakukan menggunakan metode *Knowledge Discovery in Database* (KDD) dengan algoritma C4.5 untuk mengidentifikasi beberapa faktor yang terlibat terhadap tingkat pembatalan reservasi hotel. Pengumpulan data dilakukan dengan menggunakan dataset dari situs *Kaggle* [<https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand?resource=download>]. Data yang digunakan mencakup 119.390 data observasi dari dua jenis hotel, yaitu City Hotel dan Resort Hotel, dengan rentang waktu 1 Juli 2015 hingga 31 Agustus 2017. Pada tabel 1 menunjukkan deskripsi atribut dari dataset yang dipakai untuk penelitian ini.

Tabel 1. Deskripsi atribut dataset

No	Atribut	Keterangan
1	hotel	Pembeda City Hotel dan Resort Hotel
2	is_canceled	Status pembatalan (0 = tidak, 1 = dibatalkan)
3	lead_time	Jumlah hari antara pemesanan dan check-in
4	arrival_date_year	Tahun kedatangan
5	arrival_date_month	Bulan kedatangan
6	arrival_date_week_number	Minggu keberapa dalam tahun
7	arrival_date_day_of_month	Tanggal kedatangan
8	stays_in_weekend_nights	Jumlah malam akhir pekan yang dipesan
9	stays_in_week_nights	Jumlah malam hari kerja yang dipesan
10	Adults	Jumlah tamu dewasa
11	Children	Jumlah tamu anak-anak
12	Babies	Jumlah bayi

13	Meal	Tipe makanan yang dipilih (BB, HB, FB, SC)
14	Country	Negara asal pelanggan
15	market_segment	Segmen pasar pemesanan (Online TA, Offline TA, Corporate, dll.)
16	distribution_channel	Saluran distribusi pemesanan (Direct, TA/TO, dll.)
17	is_repeated_guest	Apakah tamu adalah tamu berulang
18	previous_cancellations	Jumlah pembatalan sebelumnya oleh tamu
19	previous_bookings_not_canceled	Jumlah reservasi sebelumnya yang tidak dibatalkan
20	reserved_room_type	Tipe kamar yang dipesan
21	assigned_room_type	Tipe kamar yang diberikan
22	booking_changes	Jumlah perubahan dalam pemesanan
23	deposit_type	Jenis deposit (No Deposit, Refundable, Non Refund)
24	Agent	ID agen pemesanan (jika ada)
25	Company	ID perusahaan pemesan (jika ada)
26	days_in_waiting_list	Hari dalam daftar tunggu
27	customer_type	Tipe pelanggan (Individu, Grup, Kontrak, Transien)
28	Adr	Average Daily Rate (tarif harian rata-rata)
29	required_car_parking_spaces	Jumlah ruang parkir yang diminta
30	total_of_special_requests	Jumlah permintaan khusus
31	booking_date	Tanggal pemesanan
32	check_in / check_out date	Tanggal check-in dan check-out

1. Selection

Pada tahap ini, dilakukan pemilihan pada atribut (kolom) yang akan digunakan untuk penelitian ini. Seleksi dilakukan berdasarkan beberapa kriteria yang sudah dijelaskan pada metode penelitian. Atribut yang terpilih untuk digunakan pada penelitian ini terdapat 9 atribut yaitu *hotel type*, *booking date*, *check-in date*, *check-out date*, *is canceled*, *lead time*, *deposit type*, *customer type*, *total of special request*. Atribut yang digunakan ditunjukkan pada tabel 2.

Tabel 2. Atribut yang dipilih

No	Atribut	Keterangan
1	Hotel Type	Pembeda City Hotel dan Resort Hotel
2	Booking Date	Tanggal pemesanan
3	Check-in Date	Tanggal kedatangan
4	Check-out Date	Tanggal keberangkatan
5	Is Canceled	Informasi pembatalan
6	Lead Time	Jumlah hari antara waktu pemesanan dan waktu check-in
7	Deposit Type	Jenis deposit yang diminta oleh hotel
8	Customer Type	Klasifikasi pelanggan
9	Total of Special Request	Permintaan khusus

2. Preprocessing

Pada tahap ini, dilakukannya data *preprocessing* untuk membersihkan beberapa data yang kotor yang ada pada dataset. Hasil dari tahapan ini adalah sebuah dataset mentah yang sudah bersih dari *duplication*, *inconsistent*, dan *missing* data. Pada tahap data *preprocessing* ini hal yang dilakukan yaitu membersihkan dataset dari duplikasi yang ada dengan cara menghapus duplikasi. Selain itu, pada tahap ini juga melakukan pengisian *value* pada dataset yang kosong dengan nilai median jika kolom tersebut memiliki nilai numerik, lalu mengisi juga nilai 0 pada *missing value* atau kolom yang kosong seperti pada kolom *childrens*, *adults*, dan lainnya. Pembersihan data dilakukan dengan cara:

- Mengisi atau menghapus *missing values*
- Menghapus data duplikat
- Normalisasi *variabel*
- Mengatasi *outliers*
- Mengelompokkan *Lead Time* ke dalam tiga kategori: "Pendek" (<30 hari), "Sedang" (30-90 hari), dan "Panjang" (>90 hari)

3. Transformation

Tahap ini adalah proses merubah data kompleks menjadi sederhana agar mudah dalam proses pengelolaan. Pada tahap sebelumnya, terdapat data kategorikal yang perlu diubah kebentuk numerik untuk memudahkan proses selanjutnya. Beberapa atribut dikonversi menjadi kategori tertentu untuk meningkatkan efisiensi model klasifikasi:

- Lead Time* dikategorikan menjadi "Pendek", "Sedang", dan "Panjang".
- Jenis Deposit dikategorikan menjadi "No Deposit", "Non Refund", dan "Refundable".
- Tipe Pelanggan dikategorikan menjadi "Individu", "Bisnis", dan "Grup".

4. Data Mining

Pada tahap data mining, strategi pengujian menggunakan dataset pemesanan hotel yang telah diproses sebelumnya. Dataset ini terdiri dari 108.258 entri dengan 57 atribut setelah dilakukan preprocessing dan transformasi data. Untuk mengoptimalkan evaluasi model, dataset dibagi menjadi tiga skenario, yaitu Skenario A (90% *training*, 10% *testing*), Skenario B (80% *training*, 20% *testing*), dan Skenario C (70% *training*, 30% *testing*).

a. Penerapan Algoritma C4.5

Dalam penelitian ini, algoritma C4.5 diterapkan untuk membangun model klasifikasi pembatalan reservasi dengan tujuan untuk menganalisis dan memprediksi kemungkinan pembatalan berdasarkan berbagai faktor yang tersedia dalam dataset. Adapun tiga skenario pembagian data yang digunakan adalah sebagai berikut:

- Skenario A: 90% data digunakan untuk pelatihan (*training*), sementara 10% data digunakan untuk pengujian (*testing*).
- Skenario B: 80% data digunakan untuk pelatihan (*training*), sementara 20% data digunakan untuk pengujian (*testing*).
- Skenario C: 70% data digunakan untuk pelatihan (*training*), sementara 30% data digunakan untuk pengujian (*testing*).

Setelah membangun model klasifikasi berdasarkan masing-masing skenario, evaluasi dilakukan dengan menggunakan *confusion matrix*. Berikut adalah hasil pengujian pada setiap skenario:

1) Skenario A

Pada skenario ini, mayoritas data digunakan untuk proses pelatihan, sehingga model memiliki lebih banyak informasi untuk membangun pola klasifikasi. Hasil evaluasi dengan *confusion matrix* terdokumentasi seperti pada Gambar 2.

		Skenario A	
Actual Label	0	True Neg 26	False Pos 1
	1	False Neg 1	True Pos 23
		0	1
		Predicted Label	

Gambar 2. *Confusion matrix* pada skenario A

Berdasarkan Gambar 2, didapatkan hasil seperti:

- True Negative* (TN): 26
- False Positive* (FP): 1

- c) *False Negative* (FN): 1
- d) *True Positive* (TP): 23
- e) Akurasi: 96.08%

Akurasi yang diperoleh dalam skenario ini cukup tinggi, menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat kesalahan yang sangat kecil. Hal ini menunjukkan bahwa semakin besar proporsi data latih, semakin baik model dalam mengenali pola data.

Berdasarkan *confusion matrix* tersebut, dapat dihitung beberapa matrik evaluasi performa skenario klasifikasi, seperti pada Gambar 3.

```
===== Skenario A =====
```

	precision	recall	f1-score	support
Negative	0.962963	0.962963	0.962963	27.000000
Positive	0.958333	0.958333	0.958333	24.000000
accuracy	0.960784	0.960784	0.960784	0.960784
macro avg	0.960648	0.960648	0.960648	51.000000
weighted avg	0.960784	0.960784	0.960784	51.000000

Gambar 3. *Clasification Report* Skenario A

2) Skenario B

Dalam skenario kedua, jumlah data latih dikurangi menjadi 80%, sementara data uji ditingkatkan menjadi 20%. Tujuan dari skenario ini adalah untuk mengamati bagaimana performa model ketika proporsi data uji lebih besar dibandingkan skenario sebelumnya. Hasil evaluasi dengan *confusion matrix* terdokumentasi seperti pada Gambar 4.

Skenario B		
Actual Label	0	1
	True Neg 54	False Pos 3
1	False Neg 2	True Pos 43
		Predicted Label
		0 1

Gambar 4. *Confusion matrix* pada skenario B

Berdasarkan Gambar 4, didapatkan hasil seperti:

- a) *True Negative* (TN): 54
- b) *False Positive* (FP): 3
- c) *False Negative* (FN): 2
- d) *True Positive* (TP): 43
- e) Akurasi: 94.12%

Akurasi yang diperoleh sedikit menurun dibandingkan dengan skenario pertama, namun tetap berada dalam kategori tinggi. Hal ini menunjukkan bahwa meskipun jumlah data latih dikurangi, model masih mampu melakukan klasifikasi dengan cukup baik.

Berdasarkan *confusion matrix* tersebut, dapat dihitung beberapa matrik evaluasi performa skenario klasifikasi, seperti pada Gambar 5.

===== Skenario B =====				
	precision	recall	f1-score	support
Negative	0.964286	0.947368	0.955752	57.00000
Positive	0.934783	0.955556	0.945055	45.00000
accuracy	0.950980	0.950980	0.950980	0.95098
macro avg	0.949534	0.951462	0.950404	102.00000
weighted avg	0.951270	0.950980	0.951033	102.00000

Gambar 5. Classification Report skenario B

3) Skenario C

Pada skenario ini, proporsi data latih dikurangi lebih lanjut menjadi 70%, sementara data uji meningkat menjadi 30%. Skenario ini menguji sejauh mana model tetap dapat berfungsi dengan baik meskipun jumlah data latih semakin berkurang. Hasil evaluasi dengan *confusion matrix* terdokumentasi seperti pada Gambar 6.

		Skenario C	
Actual Label	0	True Neg 78	False Pos 8
	1	False Neg 4	True Pos 65
		0	1
		Predicted Label	

Gambar 6. Confusion matrix pada skenario C

Berdasarkan Gambar 6, didapatkan hasil seperti:

- True Negative (TN): 78
- False Positive (FP): 6
- False Negative (FN): 4
- True Positive (TP): 65
- Akurasi: 92.16%

Akurasi model mengalami sedikit penurunan dibandingkan dengan dua skenario sebelumnya, tetapi tetap menunjukkan performa yang baik. Penurunan akurasi ini dapat disebabkan oleh berkurangnya jumlah data latih, yang berpotensi menyebabkan model tidak dapat mengenali pola dalam data dengan tingkat keakuratan yang sama seperti pada skenario sebelumnya.

Berdasarkan *confusion matrix* tersebut, dapat dihitung beberapa matrik evaluasi performa skenario klasifikasi, seperti pada Gambar 7.

===== Skenario C =====				
	precision	recall	f1-score	support
Negative	0.951220	0.928571	0.939759	84.000000
Positive	0.915493	0.942029	0.928571	69.000000
accuracy	0.934641	0.934641	0.934641	0.934641
macro avg	0.933356	0.935300	0.934165	153.000000
weighted avg	0.935108	0.934641	0.934714	153.000000

Gambar 7. Classification Report skenario C

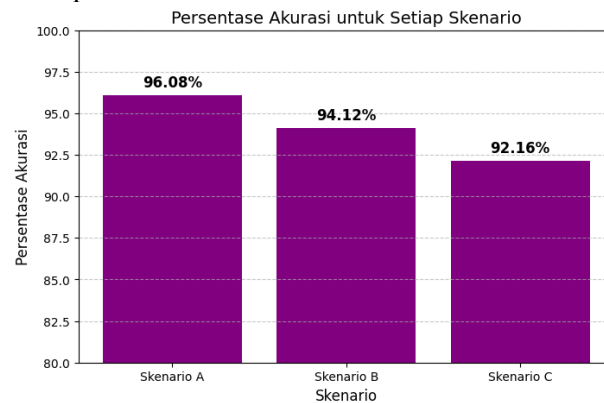
b. Evaluation

Proses evaluasi bertujuan untuk mengukur kinerja dan kualitas masing-masing skenario dalam mengklasifikasikan atau memprediksi data. Dengan menganalisis serta membandingkan tingkat akurasi dari berbagai skenario yang diterapkan, dapat ditentukan

skenario yang memiliki performa terbaik serta paling sesuai dengan tujuan penelitian. Penilaian dilakukan dengan mengevaluasi hasil prediksi dari setiap skenario.

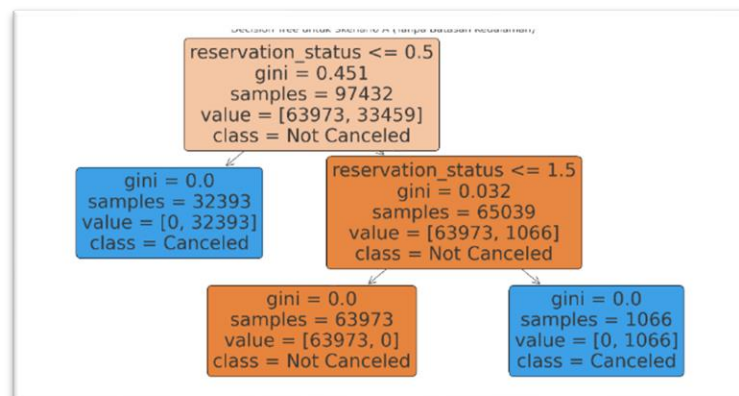
c. Akurasi Skenario

Evaluasi skenario bertujuan untuk menilai performa model dalam mengklasifikasikan data dengan membandingkan tingkat akurasi dari berbagai skenario yang telah diterapkan. Berdasarkan hasil pengujian, diperoleh persentase akurasi untuk masing-masing skenario sebagaimana ditampilkan pada Gambar 8.



Gambar 8. Persentase akurasi setiap skenario

Dari Gambar 8 diketahui nilai skenario dengan tingkat akurasi tertinggi yaitu 96.08% pada skenario A. Skenario A memberikan hasil terbaik (akurasi 96.08%) karena proporsi data latih yang lebih besar (90%) memberikan model kesempatan lebih baik dalam mempelajari pola data. Semakin banyak data yang tersedia dalam proses pelatihan, semakin akurat model dalam membentuk aturan klasifikasi. Sebaliknya, penurunan akurasi pada skenario B dan C mengindikasikan bahwa pengurangan data latih berdampak pada kemampuan model mengenali pola. Oleh karena itu, demikian *decision tree* dibuat menggunakan skenario A. *Decision Tree* dapat dilihat pada Gambar 9.



Gambar 9. Decision Tree skenario A

d. K-Fold Cross Validation

Untuk mengecek hasil performa skenario A yang terbaik sebelumnya, maka akan dilakukan evaluasi melalui *Fold Cross Validation*, seperti yang ditunjukkan dalam Tabel 3.

Tabel 3. Hasil K-Fold Cross Validation

Atribut	Keterangan
Fold 1	98.94%
Fold 2	98.87%
Fold 3	98.83%
Fold 4	98.93%
Fold 5	98.91%

Fold 6	98.88%
Fold 7	98.87%
Fold 8	98.97%
Fold 9	98.78%
Fold 10	98.97%
Rata-rata	98.90%

Dari hasil di atas, dapat disimpulkan bahwa model *Fold Cross Validation* menunjukkan performa yang stabil dan akurat dalam memprediksi pembatalan reservasi, dengan rata-rata akurasi sebesar 98.90%. Nilai ini menandakan bahwa model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya.

5. Pembahasan

Hasil penelitian menunjukkan bahwa terdapat tiga faktor utama yang mempengaruhi pembatalan reservasi hotel:

1. *Lead Time* yang Panjang

Semakin lama jarak antara waktu pemesanan dan tanggal *check-in*, semakin tinggi kemungkinan pembatalan. Hal ini disebabkan oleh meningkatnya ketidakpastian pelanggan, seperti perubahan rencana atau menemukan opsi hotel lain yang lebih menarik.

2. Jenis Deposit "No Deposit".

Pelanggan yang memilih opsi tanpa deposit lebih cenderung membatalkan reservasi karena tidak ada komitmen finansial. Sebaliknya, reservasi dengan deposit lebih kecil kemungkinan dibatalkan karena pelanggan sudah membayar sebagian dari biaya menginap.

3. Tipe Pelanggan Individu

Pelanggan individu lebih sering membatalkan reservasi dibandingkan kelompok atau pemesanan bisnis. Hal ini karena perjalanan individu lebih fleksibel dibandingkan grup atau perusahaan yang memiliki jadwal tetap.

Setelah mengetahui faktor-faktor yang berpengaruh dalam pembatalan pemesanan hotel, maka dapat ditentukan juga strategi untuk mengurangi pembatalan tersebut. Berdasarkan temuan ini, *manajemen* hotel dapat menerapkan strategi berikut:

1. Menerapkan kebijakan deposit yaitu meminta deposit saat pemesanan atau menawarkan harga lebih murah untuk reservasi yang tidak dapat dikembalikan. Beberapa jaringan hotel seperti Marriott dan Hilton telah berhasil menurunkan tingkat pembatalan dengan menetapkan kebijakan "non-refundable" untuk tarif promo.
2. Memberikan insentif untuk pemesanan cepat seperti menawarkan promo atau diskon untuk reservasi dengan *lead time* pendek guna mengurangi risiko pembatalan. Sebagai contoh yaitu Airbnb yang menawarkan diskon *early-bird* bagi pengguna yang memesan jauh-jauh hari.
3. Meningkatkan komunikasi dengan pelanggan individu dengan mengirimkan pengingat reservasi dan menawarkan fleksibilitas perubahan tanggal agar pelanggan tetap melanjutkan pemesanan. Hotel budget seperti OYO menggunakan pengingat otomatis melalui WhatsApp dan email untuk mengurangi pembatalan menit terakhir.

Dengan strategi ini, hotel dapat mengurangi pembatalan reservasi, meningkatkan okupansi kamar, dan memastikan pendapatan lebih stabil.

SIMPULAN

Penelitian ini mengidentifikasi tiga faktor utama yang memengaruhi pembatalan reservasi pada resort hotel, yaitu *lead time* yang panjang, jenis deposit "No Deposit", dan tipe pelanggan individu. Dengan menerapkan algoritma C4.5, model klasifikasi yang dibangun mampu mencapai akurasi tinggi sebesar 96,08% dan validasi rata-rata 98,90%. Temuan ini memiliki kontribusi penting baik secara praktis maupun teoritis. Secara praktis, hasil penelitian dapat dijadikan dasar bagi

manajemen hotel dalam merancang strategi pengurangan pembatalan melalui kebijakan deposit, insentif pemesanan cepat, dan peningkatan komunikasi dengan pelanggan individu. Secara teoritis, penelitian ini memperkaya literatur dalam bidang data mining dan klasifikasi dengan membuktikan efektivitas algoritma C4.5 dalam konteks industri perhotelan.

DAFTAR PUSTAKA

- Girsang, R., Ginting, E. F., & Hutasuhut, M. (2022). Penerapan Algoritma C4.5 Pada Penentuan Penerima Program Bantuan Pemerintah Daerah. *Jurnal Sistem Informasi TGD*, 1, 449–459. <https://ojs.trigunadharma.ac.id/index.php/jsi>
- Julia, M., & Masyrurroh, A. J. (2022). *LITERATURE REVIEW DETERMINASI STRUKTUR ORGANISASI: TEKNOLOGI, LINGKUNGAN DAN STRATEGI ORGANISASI*. 3(4), 383–395.
- Millah, A. S., Apriyani, Arobiah, D., Febriani, E. S., & Ramdhani, E. (2023). Analisis Data dalam Penelitian Tindakan Kelas. *Jurnal Kreativitas Mahasiswa*, 1(2), 140–153.
- Nazifah, N., & Prianto, C. (2023). Analisis Perbandingan Decision Tree Algoritma C4.5 dengan algoritma lainnya: Systematic Literature Review. *Jurnal Informatika Dan Teknologi Komputer*, 4(2), 57–64. <https://ejurnalunsam.id/index.php/jicom/>
- Nugraha, A., Maulana, I., Purwantoro, Irawan, A. S. Y., & Voutama, A. (2024). ANALISIS PENERAPAN ALGORITMA C4.5 DALAM PENENTUAN SISWA PENERIMA BEASISWA KARAWANG CERDAS (STUDI KASUS: SMK PGRI CIKAMPEK). *Jurnal Mahasiswa Teknik Informatika*, 8(5), 10078–10086.
- Prasetyo, A., Musyaffa, N., & Sastra, R. (2020). IMPLEMENTASI DATA MINING UNTUK ANALISIS DATA PENJUALAN DENGAN MENGGUNAKAN ALGORITMA APRIORI (STUDI KASUS DAPOERIN'S) Program Studi Sistem Informasi [2]. *JURNAL KHATULISTIWA INFORMATIKA*, VIII(2), 94–96.
- Purwanto, A., Primajaya, A., & Voutama, A. (2020). Penerapan Algoritma C4.5 Dalam Prediksi Potensi Tingkat Kasus Pneumonia Di Kabupaten Karawang. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 8(4), 390–396. <https://doi.org/10.26418/justin.v8i4.41959>
- Septiani, Y., Arribe, E., & Diansyah, R. (2020). ANALISIS KUALITAS LAYANAN SISTEM INFORMASI AKADEMIK UNIVERSITAS ABDURRAB TERHADAP KEPUASAN PENGGUNA MENGGUNAKAN METODE SEVQUAL (Studi Kasus : Mahasiswa Universitas Abdurrah Pekanbaru). *JURNAL TEKNOLOGI DAN OPEN SOURCE*, 3(1), 131–143.
- Syafrimansyah. (2024). Strategi Manajemen Perhotelan untuk Meningkatkan Kualitas Layanan. *AL MIKRAJ Jurnal Studi Islam Dan Humaniora*, 4(2), 1319–1325.
- Teguh, M., & Ciawati, S. T. (2020). PERANCANGAN STRATEGI DIGITAL MARKETING COMMUNICATION BAGI INDUSTRI PERHOTELAN DALAM MENJAWAB TANTANGAN ERA POSMODEREN. *Jurnal Magister Ilmu Komunikasi*, 6(1), 51–134. <http://journal.ubm.ac.id/>
- Veronica Agustin, A., & Voutama, A. (2023). IMPLEMENTASI DATA MINING KLASIFIKASI PENYAKIT DIABETES PADA PEREMPUAN MENGGUNAKAN NAÏVE BAYES. *Jurnal Mahasiswa Teknik Informatika*, 7(2), 1002–1007.
- Wisnawa, I. M. B. (2024). Analisis Bibliometrik dengan VOSViewer pada Tren Penelitian Industri Hospitality : Studi Kasus Bidang Pemasaran Hotel. *Jurnal Bisnis Hospitaliti*, 13(1), 43–62. <https://doi.org/10.52352/jbh.v13i1.1371>