

Optimalisasi Algoritma Random Forest Menggunakan SMOTE untuk Prediksi Pembatalan Tamu Hotel

Candra Agustina^{1*}, Eka Rahmawati²

¹ Sistem Informasi Akuntansi Kampus Kota Surakarta, Universitas Bina Sarana Informatika
Indonesia

² Sistem Informasi Kampus Kota Surakarta, Universitas Bina Sarana Informatika
Indonesia

* Corresponding Author. E-mail: candra.caa@bsi.ac.id)

Abstract

The role of technology has simplified user access to various services, including hotel reservations. Technology has also made it easier for service providers to offer their products by showcasing hotel photos and specifications. However, the ease of making reservations aligns with the ease for users to cancel their hotel room bookings, which undoubtedly has a negative impact on hotel owners. To minimize losses, cancellation penalties and predictions regarding the likelihood of reservation cancellations can be implemented. Therefore, predicting hotel reservation cancellations has become crucial for hotel managers to reduce occupancy declines. One of the tools for prediction can be done through machine learning technology. The researcher utilized a hotel reservation dataset consisting of 18 attributes, which underwent data processing using the Random Forest algorithm, where the label attribute was the decision to cancel or not. In this study, the Random Forest algorithm was used without optimization, yielding an accuracy of 87.85%. Subsequently, the dataset, before being processed using Random Forest, underwent handling for imbalanced data using SMOTE. Through this method, the accuracy was enhanced by 2.46%, resulting in a final accuracy of 90.31%. Since the classification was only divided into 2 groups, SMOTE was performed only once to obtain a balanced data sample. This research contributes significantly to understanding the use of machine learning algorithms, particularly Random Forest, in addressing the challenges of imbalanced data in the case of predicting hotel reservation cancellations. The research highlights the potential of SMOTE as an effective tool in handling data imbalance, which can assist hotel managers and hospitality service providers in optimizing reservation management and improving room occupancy rates.

Keywords: SMOTE, Prediction Of Cancelation, Imbalanced Data, Random Forest

Abstrak

Peran teknologi telah mempermudah pengguna untuk mengakses berbagai hal salah satunya adalah reservasi hotel. Teknologi juga mempermudah penyedia jasa untuk menawarkan produk dengan menampilkan foto dan spesifikasi dari hotel. Namun kemudahan reservasi juga selaras dengan kemudahan pengguna untuk melakukan pembatalan pemesanan kamar hotel. Hal itu tentu memberikan dampak yang buruk terhadap pemilik hotel. Untuk meminimalisir kerugian, dapat diterapkan pinalti pembatalan dan prediksi terhadap kemungkinan pembatalan reservasi. Oleh karena itu prediksi pembatalan reservasi hotel menjadi hal yang penting bagi manajer hotel untuk mengurangi penurunan occupansi. Salah satu alat untuk memprediksi dapat dilakukan

dengan teknologi machine learning. Dataset hotel reservation peneliti menggunakan 18 atribut, yang kemudian dilakukan pengolahan data dengan algoritma Random Forest, berlaku sebagai atribut label adalah keputusan untuk batal atau tidak batal. Dalam penelitian ini kami menggunakan algoritma Random Forest tanpa algoritma optimasi, hasil yang diperoleh adalah akurasi sebesar 87,85%. Selanjutnya dataset sebelum diolah menggunakan Random Forest terlebih dahulu dilakukan penanganan pada data yang tidak seimbang menggunakan SMOTE. Dengan metode ini mampu meningkatkan akurasi sebesar 2,46%, dimana hasil akhirnya adalah 90,31%. Karena klasifikasi hanya terbagi ke dalam 2 kelompok maka SMOTE cukup dilakukan 1 kali untuk mendapatkan data sample yang seimbang. Penelitian ini memberikan kontribusi penting terhadap pemahaman tentang penggunaan algoritma machine learning, khususnya Random Forest, dalam menghadapi tantangan data tidak seimbang pada kasus prediksi pembatalan reservasi hotel. Hasil penelitian menyoroti potensi SMOTE sebagai alat yang efektif dalam menangani ketidakseimbangan data, yang dapat membantu manajer hotel dan penyedia layanan perhotelan untuk mengoptimalkan pengelolaan reservasi dan meningkatkan tingkat penghunian kamar.

Kata Kunci: SMOTE; Prediksi Pembatalan Reservasi Hotel, Data Tidak Seimbang, Random Forest

1. Pendahuluan

Perkembangan industri perhotelan tidak lepas dari peran teknologi informasi. Berbagai platform untuk reservasi sangat memudahkan user untuk melakukan pemesanan hotel/penginapan. Akan tetapi kemudahan tersebut membawa beberapa dampak negatif bagi pelaku usaha perhotelan, salah satunya adalah pembatalan pemesanan hotel. Hal tersebut sangat merugikan pihak hotel, karena dengan adanya pembatalan maka pihak hotel batal mendapatkan tamu dan harus memasarkan kembali kamar yang batal di pesan. Untuk menangani hal tersebut perlu diterapkan kebijakan terkait pembatalan reservasi, karena dapat mempengaruhi occupansi sebuah hotel. Salah satu contoh adalah dengan pemberian Voucher menginap pada waktu yang akan datang. Menurut Wardani,

cara tersebut terbukti efektif untuk mengurangi kerugian bagi pihak hotel dan tamu(Wardani, 20221). Untuk mengambil kebijakan tersebut maka manajer hotel harus mempunyai Sistem Pendukung Keputusan. Kesalahan menerapkan kebijakan dapat merusak reputasi dan juga kinerja bisnis(Azhar dkk., 2021) Salah satu Sistem yang dapat diterapkan adalah dengan memprediksi pembatalan pemesanan hotel. Prediksi pembatalan sudah lazim digunakan dalam industri jasa, khususnya industri perhotelan(Antonio dkk., 2017). Setiap pembatalan menginap oleh tamu akan dikenai Cancellation Pinalty, sesuai kesepakatan sewaktu melakukan reservasi(Selamet & Sumadi, 2021). Dengan diketahui lebih awal maka kebijakan dapat diterapkan lebih awal juga sehingga sebelum melakukan konfirmasi pemesanan

tamu sudah mengetahui kebijakan pembatalan dan dapat memutuskan akan melanjutkan reservasi atau tidak.

Salah satu cara yang dilakukan adalah memanfaatkan machine learning. Machine Learning merupakan algoritma matematika yang bekerja dengan cara membaca data dan memprediksi di masa yang akan datang (Roihan dkk., 2019). Dalam beberapa yang sudah dilakukan penggunaan machine learning ini sudah terbukti dapat membantu manajer untuk menentukan strategi pengelolaan hotel. Beberapa penelitian sudah dilakukan namun dari hasil yang didapatkan masih bisa untuk ditingkatkan lagi akurasi. Semakin tinggi akurasi prediksi pembatalan tamu hotel maka akan semakin meminimalkan resiko kerugian yang dihadapi.

Sistem reservasi hotel, mempunyai beberapa resiko salah satunya adalah pembatalan. Terdapat beberapa faktor yang menyebabkan terjadinya pembatalan, salah satunya adalah keamanan, sebagai contoh jika terjadi kerusuhan, ataupun tidak kriminal dalam suatu wilayah. Selain itu pembatalan dapat diakibatkan tamu mendapatkan hotel lain dengan harga yang lebih murah (Suastini, 2018). Pembatalan pemesanan ini penting untuk dikelola dengan tujuan untuk menentukan cancellation policy atau kebijakan pembatalan. Dua hal utama yang

melatarbelakangi diperlukannya kebijakan pembatalan yaitu:

1. Meminta pertanggungjawaban kepada tamu, dan memastikan tamu memang serius hendak menginap
2. Mengurangi tamu yang batal datang, karena dapat mengurangi pendapatan (Colcol, 2023).

Dalam *Machine Learning* dikenal 4 tipe, yaitu *supervised learning*, algoritma ini memerlukan dataset yang mempunyai variabel berfungsi sebagai label. Label (*Output*) tergantung dari variabel (*Input*) pada sekelompok data training. Sedangkan tipe *Unsupervised Learning* tidak memiliki label dan tidak memerlukan data training. Algoritma ini biasanya digunakan untuk melakukan pengelompokan data berdasarkan kriteria tertentu. *Semi-Supervised Learning* menggabungkan 2 tipe sebelumnya, menggunakan sejumlah data berlabel dan data tidak berlabel untuk melakukan pengelompokan. *Machine Learning* tipe terakhir adalah *Reinforcement learning*, tipe ini memungkinkan untuk memberikan tanggapan pada setiap data. Selama proses, algoritma dapat merubah strateginya untuk mencapai nilai tertinggi (Hartanti dkk., 2023).

Implementasi *Machine Learning* dibidang perhotelan telah diterapkan pada penelitian sebelumnya dengan menggunakan berbagai algoritma untuk

melakukan analisis sentimen tamu Hotel(Puh & Bagić Babac, 2023). Penggunaan tools tersebut dapat digunakan untuk memprediksi jumlah tamu yang akan datang(Syahfitri dkk., 2020). Hasilnya akurasi prediksi sebesar 96% menggunakan algoritma backpropagation.

Penggunaan algoritma *FP Growth* dilakukan peneliti(Jafar & Rahaningsih, 2023) untuk mengetahui pola reservasi Hotel. Hasil dari pengolahan data terbentuk 5 item set dengan 25 *Assosiation Rules*. Penelitian tersebut berguna bagi manajer hotel untuk menentukan strategi marketing yang lebih efektif. Seperti contoh sebagai Sistem Pendukung Keputusan untuk menetapkan harga kamar sesuai tipe, harga paket, dan promo-promo lainnya.

Untuk memprediksi pembatalan pesanan tamu hotel, tipe yang digunakan adalah *Supervised Learning*, karena output yang dihasilkan adalah batal dan tidak batal. Penelitian sebelumnya mengemukakan bahwa optimasi *hiperparameter* dapat meningkatkan kinerja algoritma *Random Forest* dalam memprediksi pembatalan tamu hotel(Azhar dkk., 2021). Variabel yang digunakan terdapat pada Tabel 1.

Tabel 1. Penggunaan Variabel pada *Hotel Cancellation*[2]

Jenis Data	Variabel
Label	IsCanceled

Fitur	LeadTime, MarketSegment, ADR, Adults, Children, Babies, Meal, CustomerType, TotalOfSpecialRequests, PreviousCancellations, PreviousBookingsNotCanceled, BookingChanges, DaysInWaitingList, IsRepeatedGuest, DistributionChannel, StaysInWeekendNights, StaysIn-WeekNights
Fitur Kategori	DistributionChannel, Agent, MarketSegment, DepositType, CustomerType, Company, Meal

Variabel yang digunakan berjumlah 31, terdiri dari 3 jenis data yaitu label, fitur dan fitur kategoris. Penelitian tersebut memiliki akurasi lebih tinggi dibandingkan penelitian Saputro dan Nanang, mereka membandingkan beberapa algoritma dan madapatkan hasil bahwa *Random Forest* mendapatkan akurasi lebih tinggi yaitu sekitar 88%(Saputro & Nanang, 2021). Meskipun dengan dataset yang berbeda namun menunjukkan bahwa algoritma tersebut mampu diaplikasikan untuk kasus pembatalan tamu hotel.

Penelitian selanjutnya menggunakan algoritma *Decission Tree*, *Support Vector Machine* dan *Artificial Neural Network*(Hartanti dkk., 2023). Adapun hasilnya dapat dilihat pada Tabel 2.

Tabel 2. Akurasi dengan Algoritma *Decision Tree*, *Support Vector Machine* dan ANN

Algoritma	Akurasi
Decision Tree	85,7%
Support Vector Machine	82,71%
Artificial Neural Network	80,48%

Hasil penelitian pada Tabel 2 menunjukkan bahwa implementasi algoritma *Decision Tree* memiliki akurasi tertinggi dibandingkan 2 algoritma yang lainnya.

Algoritma *Random Forest* merupakan algoritma pengembangan dari *Decision Tree*, algoritma ini sudah mempunyai seleksi fitur sehingga mampu mengidentifikasi fitur-fitur yang memberi pengaruh kuat terhadap hasil, sehingga akurasi meningkat (Supriyadi dkk., 2020). Peningkatan akurasi dapat dilakukan dengan penanganan pada data yang imbalance/tidak seimbang. Dimana hasil satu kelompok lebih banyak atau dominan dibandingkan kelompok yang lainnya. Adapun metode dalam *resampling* ini adalah (Arisandi, 2023):

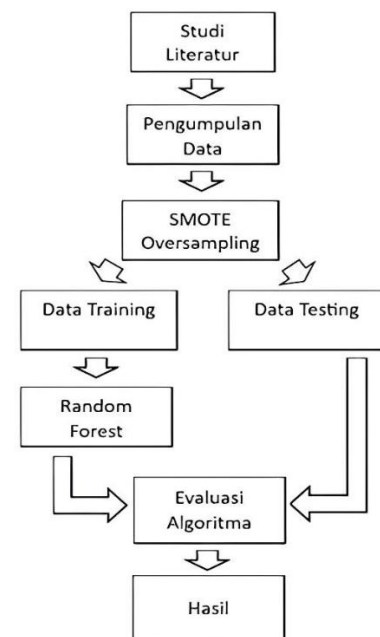
1. *Random Oversampling*, dilakukan dengan memilih kelas dengan jumlah yang sedikit secara acak kemudian ditambahkan pada data set yang baru.
2. *Random Undersampling*, metode ini dilakukan dengan memilih data pada kelas yang lebih besar kemudian dihapus sehingga jumlahnya mendekati jumlah yang minoritas.
3. *SMOTE Oversampling*, metode ini merupakan turunan dari metode *over Sampling* dilakukan dengan mensintesis kelas terkecil kemudian menduplikasinya sehingga jumlahnya bertambah (Imanwardhani, 2018).
4. *NearMiss Undersampling*, dilakukan dengan mengukur jarak sampel pada kelas mayoritas ke kelas minoritas. Sampel dengan jarak terdekat dengan kelas minoritas akan masuk kedalam kelas minoritas (Tanimoto dkk., 2022).
5. *Tomek Links*, metode *tomek links* menghapus kelas yang paling dekat dengan kelompok minoritas. Dalam

penelitian (Swana dkk., 2022), *T-Links* terbukti meningkatkan akurasi ketika dikombinasikan dengan semua algoritma.

Edited Nearest Neighbors, metode ini menghapus sampel yang jaraknya paling jauh dengan kelas mayoritas.

2. Metodologi Penelitian

Dalam rangka menggambarkan secara komprehensif proses penelitian yang dilakukan, langkah-langkah yang diimplementasikan dalam studi ini dijelaskan secara terperinci dalam diagram berikut:



Gambar 1. Metode Penelitian

Tahapan penelitian yang dilakukan adalah:

1. Studi literatur yang dilakukan bertujuan untuk mengumpulkan informasi yang relevan terkait dengan algoritma *Random Forest*, teknik *resampling*, serta topik terkait manajemen hotel, terutama terkait pembatalan reservasi hotel. Dengan melibatkan jurnal-jurnal

dan tulisan ilmiah terkini, studi literatur ini diharapkan dapat memberikan gambaran yang komprehensif tentang aplikasi dan pengaruh algoritma *Random Forest* dalam industri perhotelan, khususnya terkait pengelolaan reservasi dan prediksi pembatalan.

Dengan mempertimbangkan penelitian terdahulu dan informasi terbaru seputar topik ini, studi literatur dapat memberikan wawasan mendalam tentang pendekatan terbaik yang dapat diadopsi dalam mengoptimalkan strategi manajemen reservasi hotel. Informasi ini juga diharapkan dapat memberikan pemahaman yang lebih baik tentang bagaimana penggunaan teknik *resampling*, termasuk SMOTE, dapat membantu dalam meningkatkan akurasi prediksi dan mengatasi masalah ketidakseimbangan data. Melalui pendekatan studi literatur yang holistik dan terperinci, peneliti dapat mengevaluasi berbagai perspektif dan temuan terkait, yang pada gilirannya dapat memberikan landasan yang kuat untuk penelitian dan pengembangan lebih lanjut dalam bidang ini. Keseluruhan, studi literatur ini diharapkan dapat menjadi dasar yang kokoh bagi penelitian lebih lanjut dan implementasi praktis dalam mengoptimalkan manajemen reservasi hotel dengan memanfaatkan algoritma *Random Forest* dan teknik *resampling* yang tepat.

2. Data yang digunakan dalam penelitian ini diperoleh dari sumber terpercaya, yakni www.kaggle.com, dan telah disesuaikan kembali dalam Bahasa Indonesia untuk penamaan atributnya guna memudahkan pemahaman.

Meskipun data aslinya terdiri dari 36.276 entri, namun dalam rangka memastikan keluwesan komputasi dan analisis yang efisien, penelitian ini memilih untuk menggunakan 2.000 entri data yang mewakili sampel yang representatif. Dari total 17 atribut yang tersedia, satu di antaranya merupakan label klasifikasi, yaitu kategori "batal" atau "tidak batal" yang menjadi fokus utama dalam penelitian ini. Seluruh proses pengumpulan data dilakukan pada bulan Juli 2023 dan data tersedia dalam format *Excel*, memungkinkan peneliti untuk melaksanakan proses analisis dengan lebih mudah dan terstruktur. Dengan menggunakan dataset yang telah disaring dan disesuaikan dengan kebutuhan penelitian, diharapkan analisis yang dilakukan dapat memberikan wawasan yang berharga terkait faktor-faktor yang mempengaruhi kecenderungan pembatalan reservasi hotel, serta memungkinkan identifikasi pola atau tren yang dapat dijadikan dasar untuk pengambilan keputusan yang lebih baik di industri perhotelan.

3. *SMOTE Oversampling*, data tidak seimbang mengakibatkan hasil yang kurang akurat, oleh karena itu diterapkan *SMOTE Oversampling* untuk membuat data menjadi *balance*. Pembagian Data Training dan Data Testing, data dibagi menjadi 2 bagian yaitu untuk data training dan data testing hal ini dilakukan untuk menilai akurasi dari algoritnya yang akan digunakan. Pembagian data ini dilakukan secara acak dengan perbandingan 80:20. Pembagian data ini dilakukan menggunakan metode *K-Fold* 10 dengan perbandingan 80:20.

Teknik *K-Fold* 10 membagi dataset menjadi 10 bagian dengan perbandingan 80:20, di mana 80% dari data digunakan untuk pelatihan (*training*) dan 20% digunakan untuk pengujian (*testing*) dalam setiap iterasi. Pendekatan ini memungkinkan evaluasi yang lebih komprehensif terhadap performa model, dengan memastikan bahwa setiap bagian dataset digunakan baik untuk pelatihan maupun pengujian. Dengan mengadopsi pendekatan ini, peneliti dapat menghindari bias yang mungkin timbul dari pemilihan data yang tidak representatif. Selain itu, penggunaan metode *K-Fold* 10 memungkinkan penilaian yang lebih reliabel terhadap konsistensi performa model di berbagai subset data, yang dapat memberikan wawasan yang lebih mendalam tentang kemampuan prediktif sebenarnya dari model yang dievaluasi (Berrar, 2018). Dengan mempertimbangkan pentingnya evaluasi yang akurat, penggunaan *K-Fold* 10 dengan perbandingan 80:20 ini dapat memberikan kerangka kerja yang kokoh untuk penilaian yang cermat dan terperinci terhadap model yang sedang diuji.

4. Penerapan *Random Forest*, pengolahan data dilakukan dengan *software* Weka. Weka merupakan salah satu perangkat lunak sumber terbuka yang populer dan sering digunakan dalam analisis data dan penambahan data. Dengan antarmuka pengguna yang intuitif dan beragam alat analisis yang disediakan (Santos-Pereira dkk., 2022), Weka memungkinkan peneliti untuk melakukan pengolahan data, eksplorasi, dan pemodelan dengan cara yang efisien dan efektif. Melalui Weka,

peneliti dapat dengan mudah memuat dataset, melakukan praproses data, termasuk menangani masalah ketidakseimbangan data menggunakan teknik seperti SMOTE, serta mengimplementasikan algoritma *Random Forest* untuk memprediksi pembatalan reservasi hotel. Kemampuan Weka untuk memfasilitasi visualisasi data, evaluasi model, dan interpretasi hasil secara menyeluruh memungkinkan peneliti untuk mendapatkan wawasan mendalam tentang kinerja model serta potensi aplikasi dalam skenario praktis.

5. Evaluasi kinerja algoritma dipandang dari nilai akurasi yang dihasilkan, yang merupakan metrik penting untuk mengukur sejauh mana model dapat melakukan klasifikasi dengan benar. Akurasi memberikan gambaran tentang seberapa baik model mampu mengidentifikasi kelas data yang benar, sehingga memungkinkan perbandingan yang adil antara berbagai metode atau algoritma yang diterapkan (Power, 2020).
6. Dalam konteks ini, penggunaan akurasi sebagai metrik evaluasi memungkinkan peneliti untuk menganalisis performa algoritma *Random Forest* dalam memprediksi pembatalan reservasi hotel. Dengan mengamati nilai akurasi, peneliti dapat mengevaluasi sejauh mana model mampu mengklasifikasikan data dengan benar dan dengan demikian menilai kehandalan prediksi algoritma tersebut.

Selain itu, nilai akurasi juga memungkinkan perbandingan langsung antara metode yang berbeda, seperti teknik *resampling* atau algoritma lainnya yang mungkin diterapkan dalam konteks yang sama. Dengan

demikian, penggunaan nilai akurasi sebagai indikator evaluasi membantu memperoleh pemahaman yang lebih mendalam tentang keefektifan dan keunggulan relatif dari algoritma Random Forest dibandingkan dengan pendekatan lain yang mungkin digunakan dalam penelitian ini.

3. Hasil dan Pembahasan

Terdapat 17 atribut dan 1 label dari dataset yang digunakan. Adapun atribut terdapat pada Tabel 3.

Tabel 1. Atribut Dataset Penelitian

No	Atribut
1	no_of_adults
2	no_of_children
3	no_of_weekend_nights
4	no_of_week_nights
5	type_of_meal_plan
6	required_car_parking_spaces
7	room_type_reserved
8	lead_time
9	arrival_year
10	arrival_month
11	arrival_date
12	market_segment_type
13	repeated_guest
14	no_of_previous_cancellations
15	no_of_previous_bookings_not_canceled
16	avg_price_per_room
17	no_of_special_requests
18	booking_status

Jumlah data data yang digunakan sebanyak 2000 data. Pengujian awal dilakukan dengan mengimplemenatsikan algoritma *random forest* tanpa menggunakan *SMOTE Oversampling*. Hasil pengujian awal terdapat pada Tabel 4.

Tabel 2. Hasil Pengujian Awal dengan *Random Forest*

Keterangan	Nilai
Akurasi	87,85%
Precision	0,877
Recall	0,879
F-Measure	0,877
AUC	0,931

Akurasi dari *Random Foret* untuk data *Hotel Cancellation* mencapai 87,85%. Tabel 5 Menunjukkan *Confussion Matrix* dari pengujian awal.

Tabel 3. *Confussion Matrix* Hasil Pengujian Awal dengan *Random Forest*

	Predicted Class	
	NotCanceled	Canceled
NotCanceled	1272	97
Canceled	146	485

Tabel 5 menunjukkan terdapat 1272 *record* yang diklasifikasikan benar dan 485 diklasifikasikan salah. 146 data diprediksi benar namun seharusnya salah dan 97 data diprediksi salah namun seharusnya benar. Selanjutnya, *SMOTE Oversampling* diimplementasikan pada tahap pra-pemrosesan. Penggunaan *SMOTE Oversampling* dilandasi dari ketidakseimbangan kelas. Dari 2000 data yang digunakan, terdapat 1369 data dengan label *NotCanceled* dan 631 data dengan label *Canceled*. Setelah *SMOTE Oversampling* di aplikasikan, data dengan label *Canceled* berjumlah 1262 data sehingga total data keseluruhan adalah 2361 data. Hasil dari penerapan *SMOTE Oversampling* terdapat pada Tabel 6.

Tabel 4. Hasil Pengujian *Random Forest* dengan *SMOTE*

Keterangan	Nilai
Akurasi	90,31%
Precision	0,905
Recall	0,903
F-Measure	0,903
AUC	0,969

Penggunaan teknik *SMOTE* telah berhasil meningkatkan nilai akurasi model prediksi pembatalan reservasi hotel menjadi mencapai 90,31%, menunjukkan efektivitas metode oversampling dalam mengatasi ketidakseimbangan kelas dalam dataset. Selain peningkatan akurasi yang signifikan, hasil penelitian ini juga mengungkap adanya kenaikan nilai precision, recall, F-Measure, dan AUC (Area Under Curve). Peningkatan nilai-nilai ini mengindikasikan peningkatan signifikan dalam kemampuan model untuk mengidentifikasi dengan benar baik kelas positif maupun negatif. Dengan meningkatnya precision dan recall, model dapat memberikan prediksi yang lebih tepat dan akurat, sementara peningkatan dalam nilai F-Measure dan AUC menunjukkan performa keseluruhan model yang lebih unggul dalam mengukur keseimbangan antara presisi dan recall. Hasil *Confussion Matrix Random Forest* dengan *SMOTE* terdapat pada Tabel 7.

Tabel 5. *Confussion Matrix Random Forest* dengan *SMOTE*

	Predicted Class	
	NotCanceled	Canceled
NotCanceled	1202	167
Canceled	88 Positives	1174

Implementasi *SMOTE* dapat mengurangi data yang salah prediksi. Terdapat 167 data dengan label *NotCanceled* yang diprediksi *Canceled* dan 88 data dengan label *Canceled* yang diprediksi *NotCanceled*.

Hasil penelitian ini menunjukkan tingkat akurasi yang lebih tinggi dibandingkan dengan penelitian yang dilakukan oleh Saputro dan Nanang pada tahun 2021, di mana mereka memperoleh akurasi sebesar 79%. Perlu ditekankan bahwa perbedaan dataset yang digunakan dalam kedua penelitian tersebut dapat berkontribusi terhadap perbedaan hasil akurasi yang diamati. Namun demikian, peningkatan akurasi yang signifikan yang terlihat dalam penelitian ini, yang mencapai tingkat 90,31%, menggarisbawahi keefektifan penerapan teknik *SMOTE* Oversampling dalam mengatasi masalah ketidakseimbangan kelas dan meningkatkan performa algoritma *Random Forest* dalam memprediksi pembatalan reservasi hotel. Temuan ini menegaskan pentingnya pemilihan dataset yang tepat serta penerapan metodologi yang efektif dalam menjamin keandalan dan validitas hasil penelitian, dengan harapan dapat memberikan kontribusi berharga bagi pengembangan pengetahuan di bidang analisis prediktif dalam industri perhotelan.

Hasil ini menunjukkan bahwa teknik *SMOTE* dapat secara signifikan mengurangi kesalahan dalam memprediksi status pembatalan reservasi hotel. Dengan mengurangi jumlah kesalahan semacam itu, manajer hotel dapat mengandalkan prediksi yang lebih akurat dalam mengelola reservasi dan merencanakan strategi pemasaran yang lebih efektif. Selain itu, pengurangan kesalahan prediksi ini juga

dapat membantu meningkatkan kepuasan pelanggan dengan mengurangi ketidaknyamanan yang timbul akibat kesalahan informasi atau pemrosesan yang tidak akurat.

4. Kesimpulan

SMOTE Oversampling dapat mengoptimalkan kinerja algoritma *Random Forest* pada data *Hotel Cancellation* dengan nilai akurasi 90,31%. Tingkat akurasi mengalami kenaikan sebanyak 2,46%. Adanya peningkatan nilai akurasi menandakan *SMOTE Oversampling* dapat mengatasi masalah ketidakseimbangan kelas dengan baik. Kedepannya, dapat dilakukan optimasi dengan teknik *pre-processing* lainnya untuk meningkatkan kinerja algoritma *data mining*. Selanjutnya, peningkatan lebih lanjut dapat dilakukan dengan mengeksplorasi teknik *pre-processing* tambahan guna meningkatkan performa keseluruhan algoritma *data mining*. Integrasi metode pemrosesan data yang lebih canggih, bersama dengan algoritma pembelajaran mesin, berpotensi untuk mengembangkan model prediksi yang lebih akurat dan meningkatkan proses pengambilan keputusan dalam konteks manajemen reservasi hotel. Penelitian ini menegaskan pentingnya menggunakan teknik pemrosesan data yang canggih untuk memastikan wawasan prediktif yang akurat dan mengoptimalkan efisiensi operasional di sektor perhotelan.

References

Antonio, N., Almeida, A. de, & Nunes, L. (2017). Predicting hotel booking cancellations to decrease uncertainty and increase revenue. *Tourism & Management Studies*, 13(2), 25–39.

<https://doi.org/10.18089/tms.2017.13203>

Arisandi, R. (2023). Perbandingan Model Klasifikasi Random Forest Dengan Resampling Dan Tanpa Resampling Pada Pasien Penderita Gagal Jantung. *Jurnal Gaussian*, 12(1), 136–145. <https://doi.org/10.14710/j.gauss.12.1.136-145>

Azhar, Y., Mahesa, G. A., & Mustaqim, Moch. C. (2021). Prediction of hotel bookings cancellation using hyperparameter optimization on Random Forest algorithm. *Jurnal Teknologi dan Sistem Komputer*, 9(1), 15–21. <https://doi.org/10.14710/jtsiskom.2020.13790>

Berrar, D. (2018). Cross-validation. Dalam *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (Vol. 1–3, hlm. 542–545). Elsevier. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>

Colcol, S. (2023). *Hotel Cancellation Policy: Complete Guide*.

Hartanti, D., Pradana, A. I., & Lestari, S. (2023). *Komprasi Algoritma Decision Tree, SVM dan ANN untuk Reservasi Hotel* (Vol. 16).

Imanwardhani, C. S. (2018). Pendekatan Synthetic Minority Oversampling Technique Dalam Menangani Klasifikasi Imbalanced Data Biner.

- Jafar, J., & Rahaningsih, N. (2023). MENENTUKAN POLA RESERVASI HOTEL DENGAN ALGORITMA FP-GROWTH. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 540–546. <https://www.kaggle.com,>
- Power, D. M. W. (2020). *Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation*.
- Puh, K., & Bagić Babac, M. (2023). Predicting sentiment and rating of tourist reviews using machine learning. *Journal of Hospitality and Tourism Insights*, 6(3), 1188–1204. <https://doi.org/10.1108/JHTI-02-2022-0078>.
- Roihan, A., Abas Sunarya, P., & Rafika, A. S. (2019). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82.
- Santos-Pereira, J., Gruenwald, L., & Bernardino, J. (2022). Top data mining tools for the healthcare industry. Dalam *Journal of King Saud University - Computer and Information Sciences* (Vol. 34, Nomor 8, hlm. 4968–4982). King Saud bin Abdulaziz University. <https://doi.org/10.1016/j.jksuci.2021.06.002>
- Selamet, I. W. A., & Sumadi, I. G. (2021). PERAN RESERVASI DALAM MENINGKATKAN PENJUALAN KAMAR DI TIJILI SEMINYAK HOTEL. *Journey*, 4(2), 321–340.
- Suastini, N. M. (2018). ANALISIS MANAJEMEN RISIKO UNTUK MEMINIMALKAN DAMPAK FORCE MAJEURE DI THE RITZ-CARLTON, BALI. *Jurnal Kepariwisata*, 17(3), 5–8.
- Swana, E. F., Doorsamy, W., & Bokoro, P. (2022). Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset. *Sensors*, 22(9). <https://doi.org/10.3390/s22093246>
- Syahfitri, D., Perdana Windarto, A., & Fauzan, M. (2020). Peningkatan Nilai Akurasi Prediksi Algoritma Backpropogation (Kasus: Jumlah Pengunjung Tamu pada Hotel berbintang di Sumatera Utara). Dalam *Journal of Information System Research* (Vol. 2, Nomor 1).
- Tanimoto, A., Yamada, S., Takenouchi, T., Sugiyama, M., & Kashima, H. (2022). Improving imbalanced classification using near-miss instances. *Expert Systems with Applications*, 201. <https://doi.org/10.1016/j.eswa.2022.117130>.
- Wardani, D. M. (2021). Hotel Reservation Policy Pada Masa Pandemi : Refund, Reschedule

Atau Cancel Di Labuan Bajo.
Pariwisata, 8(1), 63–72.