

Perbandingan Teknik *Resample* pada Algoritma K-NN dan SVM untuk Prediksi Pembatalan Pemesanan Kamar Hotel

Eka Rahmawati
Program Studi *Sistem Informasi*
Universitas Bina Sarana Informatika
Jakarta, Indonesia
eka.eat@bsi.ac.id

Andria Bas Nando
Program Studi *Sistem Informasi Akuntansi*
Universitas Bina Sarana Informatika
Jakarta, Indonesia
basnando86@gmail.com

Candra Agustina*
Program Studi *Sistem Informasi Akuntansi*
Universitas Bina Sarana Informatika
Jakarta, Indonesia
candra.caa@bsi.ac.id

Fadila Chika Kusumarini
Program Studi *Sistem Informasi*
Universitas Bina Sarana Informatika
Jakarta, Indonesia
fadilachikakusumarini@gmail.com

Abstract— The increased ease of hotel reservation directly correlates with a rise in cancellation convenience, posing a potentially detrimental impact on hotel management. Consequently, a decision support system is necessary to assist managers in making informed decisions amidst the dynamic hospitality industry landscape. One approach applied is leveraging data mining technology, focusing on the K-Nearest Neighbors (kNN) and Support Vector Machine (SVM) algorithms to predict hotel room reservation cancellations. This study involves the implementation of resampling techniques to enhance prediction accuracy. The research results indicate that SVM achieves an accuracy rate of 85.05%, while SVM, supported by resampling, increases to 93.85%. On the other hand, KNN achieves an accuracy of 80.1%, and with resampling implementation, its accuracy rises to 92.15%. These findings underscore the significant potential of data mining technology, particularly in the context of predicting hotel room reservation cancellations, with a substantial increase in accuracy through the application of resample techniques to both k-NN and SVM algorithms.

Keywords—*Resample; K-NN; SVM; Hotel Room Reservation Cancellations.*

Abstrak— Peningkatan kemudahan dalam pemesanan reservasi hotel secara langsung berdampak pada peningkatan kemudahan pembatalan, memberikan dampak yang potensial merugikan bagi pengelola hotel. Oleh karena itu, diperlukan sistem pendukung keputusan yang dapat membantu para manajer dalam membuat keputusan yang cerdas di tengah dinamika industri perhotelan. Salah satu pendekatan yang diterapkan adalah memanfaatkan teknologi data mining, dengan fokus pada algoritma K-Nearest Neighbors (kNN) dan Support Vector Machine (SVM) untuk melakukan prediksi pembatalan pemesanan kamar hotel. Penelitian ini melibatkan penerapan teknik resampling guna meningkatkan akurasi prediksi. Hasil penelitian mengindikasikan bahwa SVM mencapai tingkat akurasi sebesar 85,05%, sementara SVM yang didukung oleh teknik resample meningkat menjadi 93,85%. Di sisi lain, KNN mencapai akurasi 80,1%, dan dengan penerapan resampling, akurasinya meningkat menjadi 92,15%. Temuan ini menunjukkan potensi besar teknologi data mining, khususnya dalam konteks prediksi pembatalan pemesanan

kamar hotel, dengan peningkatan signifikan akurasi melalui penerapan teknik resample pada kedua algoritma kNN dan SVM.

Keywords—*Resample; K-NN; SVM; Pembatalan Pemesanan Kamar Hotel.*

PENDAHULUAN

Pemesanan kamar hotel merupakan proses yang kompleks dalam industri perhotelan. Salah satu tantangan utama yang dihadapi oleh penyelenggara hotel adalah prediksi pembatalan pemesanan. Pembatalan pemesanan dapat dipicu oleh berbagai faktor, termasuk perubahan rencana perjalanan atau penawaran lebih menguntungkan dari hotel lain. Pengembangan model prediktif yang dapat memprediksi dengan akurat kemungkinan pembatalan pemesanan kamar menjadi krusial dalam manajemen operasional hotel[1]. Selain itu, lingkungan bisnis yang dinamis dan persaingan yang ketat dalam industri perhotelan menuntut adanya strategi yang inovatif dan responsif. Dalam konteks ini, teknologi dan analisis data menjadi kunci untuk meningkatkan efisiensi dan meningkatkan pengalaman pelanggan.

Penelitian melakukan analisis perbandingan terhadap dua algoritma pembelajaran mesin, yaitu *K-Nearest Neighbors* (K-NN) dan *Support Vector Machine* (SVM), dengan menerapkan teknik *resampling*. Teknik *resampling* merupakan pendekatan yang efektif untuk mengatasi ketidakseimbangan data, yang seringkali terjadi dalam pengolahan data[2]. Penelitian sebelumnya menunjukkan bahwa algoritma *K-Nearest Neighbors* (K-NN) mampu mencapai tingkat akurasi yang lebih tinggi dibandingkan dengan algoritma Logistic Regression. Hasil ini menarik perhatian karena menandakan kemampuan K-NN dalam menangani permasalahan tertentu yang mungkin sulit diakomodasi oleh *Logistic Regression*[3].

Kelebihan k-NN dibandingkan dengan algoritma yang lain[4]:

A. Pelatihan Sangat Cepat

Kelebihan ini terutama terlihat karena k-NN merupakan algoritma pembelajaran mesin yang termasuk dalam kategori

"*lazy learning*" atau "*instance-based learning*" yang berarti k-NN tidak benar-benar melakukan proses pelatihan pada data. Proses pelatihan hanya terjadi ketika algoritma diimplementasikan untuk membuat prediksi pada data yang belum pernah dilihat sebelumnya. Waktu pelatihan k-NN sangat cepat, karena tidak ada proses pembangunan model yang memerlukan iterasi pada data pelatihan.

B. Sederhana dan Mudah Dipelajari

K-NN adalah salah satu algoritma yang sederhana dan mudah dipahami. Konsep dasarnya adalah menghitung jarak antara data yang baru dan data pelatihan yang ada untuk menemukan tetangga terdekat. Tanpa perlu mengasumsikan distribusi tertentu atau parameter kompleks, k-NN memberikan pendekatan yang intuitif dan mudah diinterpretasikan.

C. Tahan Terhadap Data Pelatihan yang Memiliki Derau

K-NN cenderung tahan terhadap *noise* atau gangguan dalam data pelatihan. Ketika ada nilai yang tidak biasa atau data yang mencurigakan, k-NN dapat mengatasi hal ini dengan mempertimbangkan sejumlah tetangga terdekat, sehingga dampak data yang tidak biasa dapat diurutkan dengan tetangga lainnya.

D. Efektif jika Data Pelatihan Besar

K-NN dapat efektif digunakan ketika data pelatihan cukup besar. Hal tersebut dikarenakan k-NN tidak melibatkan proses pelatihan yang rumit dan tidak memerlukan penyimpanan model yang besar. K-NN hanya perlu menyimpan data pelatihan itu sendiri, dan proses prediksi dapat dilakukan dengan cepat berdasarkan perhitungan jarak.

Kelebihan-kelebihan tersebut membuat k-NN menjadi pilihan yang baik untuk aplikasi tertentu, terutama ketika kompleksitas model tidak menjadi faktor kritis dan interpretabilitas serta kecepatan pelatihan adalah pertimbangan utama. Hasil penelitian memberikan kontribusi positif terhadap pemahaman tentang pilihan algoritma pembelajaran mesin dalam menangani permasalahan prediksi. Selain itu, dapat menjadi dasar untuk mempertimbangkan penggunaan K-NN ketika dihadapkan pada dataset atau kasus yang cenderung kompleks dan tidak linear. Namun, seiring dengan keunggulan yang dimiliki oleh K-NN, perlu juga diingat bahwa pemilihan algoritma harus disesuaikan dengan sifat dan karakteristik khusus dari dataset yang digunakan.

Algoritma *Support Vector Machine* (SVM) digunakan sebagai pendekatan untuk memprediksi pembatalan reservasi hotel[5], dan hasil penelitian menunjukkan tingkat akurasi sebesar 82,71%. Hasil tersebut mengindikasikan keberhasilan SVM dalam menangani kompleksitas dan pola data terkait dengan perilaku pembatalan pemesanan kamar hotel. SVM merupakan algoritma yang kuat untuk tugas klasifikasi dan regresi, yang bekerja dengan mencari batas keputusan optimal di antara kelas-kelas yang berbeda dalam ruang fitur. Keunggulan SVM terletak pada kemampuannya menangani hubungan non-linear dan menangkap pola kompleks dalam data, terutama ketika terdapat ketidakseimbangan antara kelas-kelas. Penggunaan SVM dalam konteks prediksi pembatalan reservasi hotel menjadi signifikan karena dapat memberikan informasi yang berharga kepada penyelenggara hotel. Dengan memahami dan memprediksi perilaku pembatalan, hotel dapat mengoptimalkan manajemen reservasi, meningkatkan efisiensi operasional, dan

mengurangi dampak keuangan yang mungkin timbul akibat kamar yang tidak terpakai.

Meskipun akurasi sebesar 82,71% adalah pencapaian yang signifikan, penting untuk diperhatikan bahwa evaluasi kinerja SVM tidak hanya terbatas pada tingkat akurasi. Evaluasi yang holistik, termasuk analisis matriks kebingungan, presisi, recall, dan F1-score, dapat memberikan wawasan lebih mendalam tentang kemampuan model dalam menangani kelas minoritas, seperti kasus pembatalan pemesanan yang seringkali merupakan kejadian yang relatif jarang terjadi. Dengan demikian, penggunaan SVM dalam prediksi pembatalan reservasi hotel[5] membuka potensi untuk penerapan teknologi dalam meningkatkan efektivitas manajemen hotel, dan penelitian ini memberikan kontribusi penting terhadap pemahaman kita tentang aplikasi SVM dalam konteks industri perhotelan.

Selain mengevaluasi kinerja model berdasarkan tingkat akurasi, pada penelitian Siringoringo mengemukakan analisis terhadap presisi antara model k-NN dan R+k-NN. Seperti halnya dengan aspek akurasi, penerapan metode *resampling* pada model k-NN juga terbukti memberikan perbaikan yang signifikan terhadap kinerja klasifikasi. *Resampling*, dalam konteks ini, dapat dianggap sebagai strategi yang berhasil dalam meningkatkan ketepatan presisi prediksi k-NN, dan perbandingan dengan model R+k-NN memberikan wawasan lebih lanjut tentang efektivitas teknik ini. Proses *resampling* pada model k-NN secara esensial bertujuan untuk mengatasi ketidakseimbangan dalam data, yang sering kali menjadi tantangan dalam tugas klasifikasi. Dengan melakukan pengulangan atau pengurangan jumlah sampel pada setiap kelas, *resampling* membantu meningkatkan ketepatan prediksi, terutama dalam menangani kelas minoritas yang mungkin memiliki kontribusi penting terhadap keputusan akhir.

Hasil penelitian ini menyoroti bahwa, selain mempertimbangkan akurasi, peningkatan kinerja klasifikasi dapat diperoleh melalui strategi *resampling* pada model k-NN. Dengan memperbaiki presisi, *resampling* dapat memberikan manfaat tambahan dalam konteks analisis klasifikasi yang lebih mendalam, memberikan informasi yang lebih kaya dan akurat terkait dengan performa model. Oleh karena itu, pemahaman terhadap efek *resampling* terhadap presisi klasifikasi pada model k-NN menjadi penting dalam merinci keefektifan teknik ini dalam meningkatkan kinerja model pembelajaran mesin[6][7].

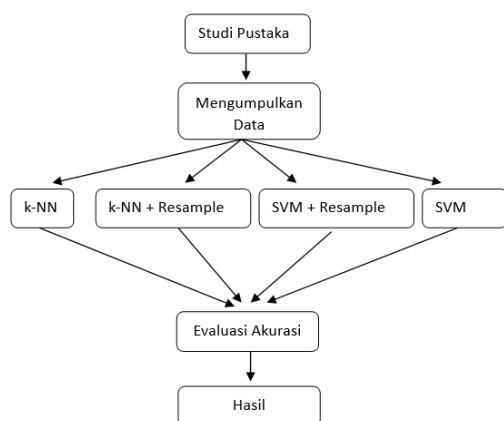
Nugraha dan Sabarudin berhasil menunjukkan bahwa penggunaan teknik *resampling* dan metode *Support Vector Machine* (SVM) memberikan hasil akurasi yang lebih tinggi dalam konteks klasifikasi penderita diabetes dibandingkan dengan penggunaan algoritma lain. Penelitian ini menyoroti efektivitas kombinasi antara teknik *resampling*, yang berfokus pada penanganan ketidakseimbangan data, dan SVM sebagai algoritma klasifikasi. Teknik *resampling* yang diterapkan dalam penelitian ini, bertujuan untuk mengatasi masalah ketidakseimbangan kelas pada data. Hal ini dapat melibatkan *oversampling*, *undersampling*, atau metode lainnya yang menghasilkan distribusi kelas yang lebih seimbang. Pendekatan ini terbukti menjadi strategi yang sukses dalam meningkatkan kinerja klasifikasi, khususnya ketika dihadapkan pada kasus-kasus di mana kejadian minoritas seperti diabetes perlu dikenali dengan lebih baik.

SVM memanfaatkan konsep *hyperplane* untuk memisahkan dua kelas dalam ruang fitur. Keunggulan SVM terletak pada kemampuannya menangani hubungan non-linear dan kompleks antara variabel input, yang seringkali muncul dalam masalah kesehatan seperti klasifikasi penderita diabetes. Dengan demikian, temuan dari penelitian Nugraha dan Sabarudin menunjukkan bahwa kombinasi teknik resampling dan SVM memberikan solusi yang efektif untuk meningkatkan akurasi[8].

Tujuan dari penelitian ini adalah mengevaluasi kinerja kedua algoritma dalam menghadapi potensi pembatalan pemesanan kamar hotel. Dengan menerapkan teknik *resampling*, penelitian mengidentifikasi pendekatan yang paling efisien untuk meningkatkan akurasi prediksi. Hal ini diharapkan dapat memberikan wawasan yang bermanfaat bagi industri perhotelan dalam memperbaiki manajemen reservasi dan optimalisasi pendapatan. Lebih lanjut, harapannya, hasil penelitian ini dapat membentuk dasar untuk pengembangan model prediktif yang lebih maju dan adaptif terhadap dinamika pasar, membantu hotel dalam membuat keputusan yang lebih cerdas, serta menghadapi tantangan prediksi pembatalan dengan lebih efektif.

METODE PENELITIAN

Metode penelitian terkait prediksi pembatalan pemesanan kamar hotel terdapat pada Gambar 1.



Gambar 1 Metode Penelitian

A. Studi Pustaka

Studi pustaka merupakan metode penelitian yang mengintegrasikan pengumpulan informasi dan data dari berbagai sumber, baik yang terdapat di perpustakaan maupun yang dapat diakses melalui internet, yang relevan dengan permasalahan yang sedang diinvestigasi. Pendekatan ini mencakup kegiatan terorganisir untuk menyelidiki, mengumpulkan, memproses, dan merangkum data dengan menerapkan metode atau teknik tertentu.

Dalam praktiknya, studi pustaka melibatkan pencarian dan analisis terhadap literatur-literatur, jurnal-jurnal ilmiah, buku-buku, dan sumber informasi lainnya yang dapat memberikan wawasan mendalam terkait dengan subjek penelitian. Proses ini dilakukan secara sistematis untuk memastikan bahwa informasi yang diperoleh dapat menjadi dasar yang kokoh dalam menghadapi permasalahan penelitian yang tengah dihadapi.

Selain itu, studi pustaka juga dapat mencakup penggunaan sumber-sumber daring atau internet, yang seringkali menyediakan akses ke informasi terkini dan pemikiran terbaru dari berbagai disiplin ilmu. Penerapan metode ini tidak hanya bertujuan untuk mengidentifikasi kerangka teoretis yang relevan, tetapi juga untuk mendapatkan pemahaman mendalam tentang perkembangan dan perdebatan yang berkaitan dengan topik penelitian.

Tujuan utama dari kegiatan studi pustaka adalah untuk mendapatkan jawaban atau solusi yang solid terhadap permasalahan yang sedang diinvestigasi. Dengan mengumpulkan dan menganalisis informasi dari berbagai sumber, peneliti dapat membentuk landasan teoritis yang kuat dan merinci pemahaman mendalam terhadap isu-isu yang terkait dengan penelitian tersebut.

Dengan demikian, studi pustaka bukan hanya menjadi langkah awal dalam perjalanan penelitian, melainkan juga menjadi fondasi yang mendukung dan memberikan arah bagi pengembangan metodologi, analisis data, dan interpretasi hasil selanjutnya. Keseluruhan proses studi pustaka berkontribusi pada kualitas dan ketepatan penelitian, membantu peneliti dalam memahami konteks lebih luas dari topik yang dipilih, serta memastikan bahwa penelitian tersebut dapat memberikan kontribusi yang berarti dalam ranah ilmiah.[9].

B. Data Penelitian

Penelitian ini dapat dikategorikan sebagai penelitian sekunder karena menggunakan data yang telah ada dan diperoleh dari kaggle.com. Dalam konteks penelitian sekunder, peneliti tidak mengumpulkan data langsung dari lapangan, melainkan memanfaatkan dataset yang sudah ada sebagai sumber informasi[10]. Keuntungan dari pendekatan penelitian sekunder ini termasuk efisiensi waktu dan biaya, karena peneliti tidak perlu melakukan pengumpulan data secara mandiri. Pemilihan dataset dari kaggle.com, sebuah platform yang dikenal menyediakan berbagai dataset dalam berbagai bidang, menambah validitas dan keandalan hasil penelitian.

Selain itu, kaggle.com juga umumnya menyertakan dokumentasi dan metadata yang membantu peneliti dalam memahami konteks data yang digunakan. Dengan demikian, penelitian ini mengedepankan aspek analisis dan interpretasi data, memanfaatkan dataset yang sudah ada untuk mendapatkan wawasan yang berharga terkait prediksi pembatalan pemesanan kamar hotel dengan memanfaatkan teknik *resampling* pada algoritma *K-Nearest Neighbors* (kNN) dan *Support Vector Machine* (SVM).

C. Evaluasi

Selain nilai akurasi, performa algoritma juga dapat dievaluasi melalui Area Under the Curve (AUC), sebagaimana ditemukan dalam penelitian yang dilakukan oleh Ardiyansyah dan rekan-rekannya. AUC merupakan metrik evaluasi yang umum digunakan dalam konteks klasifikasi, terutama pada kurva *Receiver Operating Characteristic* (ROC). Metrik ini memberikan gambaran tentang kemampuan model dalam memisahkan kelas 'Ya' dan 'Tidak'. Penelitian Ardiyansyah dkk. menunjukkan bahwa penilaian kinerja algoritma tidak hanya terbatas pada akurasi semata, tetapi juga melibatkan pengukuran AUC untuk memberikan perspektif yang lebih holistik terhadap keefektifan model. AUC dapat

memberikan informasi tambahan tentang sejauh mana model mampu membedakan antara kelas-kelas yang diuji[11].

Confusion Matrix juga digunakan dalam evaluasi. *Confusion matrix* adalah alat yang digunakan untuk mengevaluasi kinerja model klasifikasi dalam *machine learning* dengan bentuk tabel yang memungkinkan visualisasi performa model dengan membandingkan nilai aktual dari dataset dengan nilai yang diprediksi oleh model[12]. Dalam *confusion matrix*, terdapat empat istilah utama:

- *True Positive (TP)*: Jumlah sampel positif yang diprediksi dengan benar oleh model.
- *True Negative (TN)*: Jumlah sampel negatif yang diprediksi dengan benar oleh model.
- *False Positive (FP)*: Jumlah sampel negatif yang salah diprediksi sebagai positif oleh model (juga dikenal sebagai *Type I error*).
- *False Negative (FN)*: Jumlah sampel positif yang salah diprediksi sebagai negatif oleh model (juga dikenal sebagai *Type II error*).

Nilai *precision* dan *recall* juga menjadi bagian dari evaluasi[13]. *Precision* merupakan kemampuan model untuk mengidentifikasi secara akurat kelas positif ((TP) / (TP + FP)). Sedangkan *recall (Sensitivity atau True Positive Rate)* adalah kemampuan model untuk menemukan kembali semua instance dari kelas yang benar ((TP) / (TP + FN)).

HASIL DAN PEMBAHASAN

Terdapat 2000 pemesanan kamar hotel data yang diolah dengan 19 atribut. Adapun atribut yang dimiliki oleh dataset terdapat pada Tabel 1.

TABEL I. ATRIBUT DATASET PENELITIAN

| No | Nama Atribut |
|----|--------------------------------------|
| 1 | Booking_ID |
| 2 | no_of_adults |
| 3 | no_of_children |
| 4 | no_of_weekend_nights |
| 5 | no_of_week_nights |
| 6 | type_of_meal_plan |
| 7 | required_car_parking_space |
| 8 | room_type_reserved |
| 9 | lead_time |
| 10 | arrival_year |
| 11 | arrival_month |
| 12 | arrival_date |
| 13 | market_segment_type |
| 14 | repeated_guest |
| 15 | no_of_previous_cancellations |
| 16 | no_of_previous_bookings_not_canceled |
| 17 | avg_price_per_room |

| | |
|----|------------------------|
| 18 | no_of_special_requests |
| 19 | booking_status |

Dalam konteks pengujian awal, di mana fokusnya adalah pada satu label tunggal, yaitu "booking_status" yang dapat berisi informasi apakah suatu pemesanan (booking) dibatalkan atau tidak, analisis dilakukan dengan menerapkan algoritma K-NN (K-Nearest Neighbors) dan SVM (Support Vector Machine). Dengan 19 atribut yang memuat berbagai informasi terkait pemesanan, tujuan utama adalah untuk memahami dan meningkatkan kemampuan model dalam memprediksi apakah suatu booking akan mengalami pembatalan atau tidak. Hasil pengujian awal terdapat pada Tabel 2.

TABEL II. HASIL PENGUJIAN AWAL

| Algoritma | Akurasi | Precision | Recall | AUC |
|-----------|---------|-----------|--------|-------|
| K-NN | 80,1% | 0,803 | 0,801 | 0,798 |
| SVM | 85,05% | 0,848 | 0,851 | 0,815 |

Implementasi algoritma K-NN memiliki nilai akurasi 80,1% sehingga akurasi tersebut tidak lebih baik jika dibandingkan dengan algoritma SVM. Efektivitas SVM juga ditunjukkan dari hasil AUC sebesar 0,815.

Pada tahap berikutnya dari penelitian ini, penelitian diteruskan dengan mengimplementasikan teknik *resample*. Data memiliki ketidakseimbangan kelas pada label *booking status* dimana status *Not Canceled* berjumlah 1369 dan data *Canceled* berjumlah 631. *Resampling* dapat digunakan untuk mengatasi ketidakseimbangan kelas[15]. Tujuan dari penerapan teknik ini adalah untuk meningkatkan akurasi model secara substansial, di mana strategi ini melibatkan penggunaan ulang sampel-sampel data yang sudah ada. Dengan cara ini, diharapkan bahwa model yang dihasilkan akan menjadi lebih optimal dan dapat mengatasi potensi ketidakseimbangan atau keterbatasan pada data, menciptakan landasan yang lebih kuat untuk analisis dan prediksi yang akurat. Adapun hasil pengujian dengan *resample* terdapat pada Tabel 3.

TABEL III. HASIL PENGUJIAN DENGAN RESAMPLE

| Algoritma | Akurasi | Precision | Recall | AUC |
|-----------|---------|-----------|--------|-------|
| K-NN | 92,15% | 0,921 | 0,922 | 0,907 |
| SVM | 93,85% | 0,938 | 0,939 | 0,920 |

Hasil pengujian yang dilakukan menggambarkan peningkatan yang signifikan dalam kinerja algoritma, terutama pada algoritma K-NN dan SVM, setelah penerapan teknik *resample*. Peningkatan ini sangat terlihat dari peningkatan nilai akurasi dan Area Under the Curve (AUC) yang mencirikan kemampuan model dalam memisahkan antara kelas positif dan negatif.

Pertama, pada algoritma K-NN, terdapat peningkatan yang mencolok dalam nilai akurasi. Sebelum implementasi *resample*, akurasi sebesar 80,1%, namun setelahnya

meningkat drastis menjadi 92,15%. Hal ini menandakan bahwa dengan memanfaatkan teknik resample, model K-NN mampu mengenali dan mengklasifikasikan data dengan tingkat keakuratan yang jauh lebih tinggi. Selain itu, perbaikan yang terlihat tidak hanya sebatas pada akurasi, tetapi juga tercermin dalam nilai AUC yang meningkat dari 0,798 menjadi 0,907. AUC yang lebih tinggi menunjukkan bahwa model memiliki kemampuan yang lebih baik dalam membedakan antara kelas positif dan negatif, menggambarkan peningkatan keseluruhan dalam kualitas prediksi model.

Peningkatan yang sama juga diamati pada algoritma SVM. Setelah menerapkan resample, akurasi naik menjadi 93,85%, menciptakan perbaikan yang nyata dari kondisi sebelumnya. Lebih lanjut, nilai AUC yang meningkat menjadi 0,920 menunjukkan bahwa model SVM dengan teknik resample mampu meningkatkan keandalannya dalam mengklasifikasikan dan memprediksi data. Peningkatan ini dapat memberikan keyakinan tambahan dalam kegunaan model dalam situasi di mana akurasi dan kemampuan membedakan antara kelas sangat kritis.

Secara keseluruhan, hasil-hasil ini menyoroti potensi dan keefektifan penggunaan teknik resample, tidak hanya dalam meningkatkan nilai akurasi tetapi juga dalam meningkatkan kemampuan diskriminatif model, seperti yang tercermin dalam nilai AUC. Penerapan teknik ini mungkin sangat bernilai dalam kasus-kasus di mana ketepatan prediksi dan kemampuan memisahkan antara kelas positif dan negatif memiliki dampak besar terhadap keberhasilan suatu model machine learning. Selanjutnya, dapat dilihat *confussion matrix* dari pengujian awal dan pengujian dengan teknik *reample*. Hasil *confussion matrix* terdapat pada Tabel 4.

TABEL IV. CONFUSION MATRIX

| Algorithm | Values | Actual Values | | |
|---------------|------------------|---------------|---------|-----|
| | | Class=Y | Class=N | |
| KNN | Predicted Values | Class=Y | 1160 | 209 |
| | | Class=N | 189 | 442 |
| SVM | Predicted Values | Class=Y | 1248 | 121 |
| | | Class=N | 178 | 453 |
| KNN+ Resample | Predicted Values | Class=Y | 1291 | 78 |
| | | Class=N | 79 | 552 |
| SVM+ Resample | Predicted Values | Class=Y | 1328 | 41 |
| | | Class=N | 82 | 549 |

Confusion matrix menjadi sebuah instrumen penting dalam mengevaluasi performa algoritma machine learning, dan dalam konteks khusus algoritma Support Vector Machines (SVM) dengan penerapan teknik resampling, terlihat adanya kecenderungan yang sangat positif terutama dalam nilai true positive. Dalam sebuah dataset yang terdiri dari 2000 sampel, tergambar dengan jelas bahwa model SVM, yang ditingkatkan dengan penggunaan teknik resample, menunjukkan kehebatannya dalam mengenali dengan tepat instansi positif. Lebih spesifik lagi, dari total 2000 sampel

tersebut, sebanyak 1328 sampel positif berhasil diprediksi secara akurat oleh algoritma.

Angka *true positive* yang mencapai jumlah yang signifikan ini menjadi suatu indikator kuat bahwa model tidak hanya mampu mengenali, tetapi juga mengklasifikasikan dengan benar sebagian besar sampel positif[14]. Keunggulan ini menjadi krusial terutama dalam situasi di mana fokus utama adalah untuk mengidentifikasi dengan akurat instansi dari kelas tertentu, seperti pada kasus diagnosis medis atau deteksi penipuan. Penggunaan teknik resample turut memperkuat kinerja SVM, khususnya dalam penanganan dataset yang tidak seimbang, di mana jumlah satu kelas jauh lebih dominan dibanding yang lain. Dengan mengatasi ketidakseimbangan ini melalui resampling, algoritma SVM mampu dengan lebih cermat menangkap nuansa dari kelas minoritas, yang tercermin dalam peningkatan jumlah true positive.

Dalam aplikasi praktis, tingginya nilai true positive memberikan keyakinan bahwa model memiliki kemampuan rendah untuk menghasilkan false negatives, yang berarti risiko melewatkan kasus positif menjadi berkurang. Hal ini sangat bermanfaat dalam konteks di mana ketersediaan mendeteksi kasus positif memiliki dampak serius. Secara keseluruhan, integrasi SVM dengan teknik resample, sebagaimana tercermin dalam confusion matrix, menggambarkan sebuah keseimbangan yang kuat antara ketepatan dan ketangguhan model dalam memprediksi dengan akurat sampel-sampel positif, memberikan keyakinan ekstra terhadap kehandalan model untuk tugas-tugas yang memerlukan identifikasi kelas tertentu secara andal.

KESIMPULAN

Kesimpulan dari penelitian ini menegaskan bahwa penerapan algoritma K-NN dan SVM dalam konteks memprediksi pembatalan pemesanan kamar hotel berhasil mencapai tingkat akurasi yang sangat memuaskan. Penerapan teknik resample secara signifikan telah membuktikan potensinya dalam meningkatkan beberapa metrik evaluasi kunci, termasuk akurasi, precision, recall, dan Area Under the Curve (AUC) pada kedua algoritma K-NN dan SVM.

Dalam konteks perbandingan antara K-NN dan SVM, evaluasi menunjukkan bahwa SVM memiliki tingkat akurasi yang lebih tinggi daripada K-NN. Hal ini menandakan bahwa SVM memiliki keunggulan dalam mengenali dan mengklasifikasikan pola-pola yang lebih kompleks dalam data pembatalan pemesanan kamar hotel. Performa yang lebih tinggi dari SVM dapat memberikan kepercayaan tambahan terhadap kemampuannya untuk menangani situasi yang lebih rumit atau variabilitas yang lebih besar dalam data.

Namun, kesimpulan ini juga memberikan ruang untuk penelitian lebih lanjut. Dengan merinci bahwa penelitian selanjutnya dapat mengeksplorasi penggunaan teknik lainnya untuk meningkatkan nilai akurasi algoritma, memberikan landasan untuk eksperimen lebih lanjut. Dengan mengintegrasikan teknik-teknik baru atau mengkombinasikan metode-metode yang sudah ada, penelitian berikutnya dapat memperdalam pemahaman kita tentang keberagaman teknik yang dapat digunakan untuk meningkatkan kinerja algoritma dalam konteks prediksi pembatalan pemesanan kamar hotel. Dengan demikian, penelitian ini memberikan kontribusi yang berharga dan memberikan arah untuk penelitian lebih lanjut dalam pengembangan model prediksi yang lebih canggih dan andal.

UCAPAN TERIMA KASIH

Tim peneliti dengan ingin mengungkapkan rasa terima kasih yang sebesar-besarnya kepada Universitas Bina Sarana Informatika atas dukungan keuangan yang luar biasa yang telah diberikan untuk proyek penelitian ini. Selain itu, tim peneliti juga merasa terhormat dan berterima kasih kepada pihak lain yang turut memberikan dukungan kepada tim. Bantuan dan dorongan dari berbagai pihak, entah itu dalam bentuk pemikiran, masukan, atau bimbingan, memberikan warna yang beragam pada perjalanan penelitian ini. Keterlibatan mereka membuktikan bahwa penelitian bukan hanya tentang kelompok peneliti semata, melainkan kolaborasi dari individu-individu yang berkomitmen untuk mencapai hasil yang bermakna.

REFERENSI

- [1] D. M. Wardani, "Hotel Reservation Policy Pada Masa Pandemi : Refund, Reschedule Atau Cancel Di Labuan Bajo," *Pariwisata*, vol. 8, no. 1, pp. 63–72, 2021.
- [2] G. L. Pritalia, "Analisis Komparatif Algoritme Machine Learning pada Klasifikasi Kualitas Air Layak Minum," *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol. 2, no. 1, 2022.
- [3] M. F. Sholahuddin *et al.*, "Perbandingan Model Logistic Regression dan K-Nearest Neighbors Dalam Prediksi Pembatalan Hotel," *Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika*, Mar. 2023, doi: 10.31284/p.snestik.2023.4040.
- [4] S. Ulya, M. Arief Soeleman, F. Budiman, and M. Teknik Informatika, "Optimasi Parameter K Pada Algoritma K-NN Untuk Klasifikasi Prioritas Bantuan Pembangunan Desa," *Techno.COM*, vol. 20, no. 1, pp. 83–96, 2021.
- [5] D. Hartanti, A. I. Pradana, and S. Lestari, "Komparasi Algoritma Decision Tree, SVM dan ANN untuk Reservasi Hotel," *Jurnal Ilmiah Teknologi Informasi dan Komunikasi*, vol. 16, 2023.
- [6] R. Siringoringo, "Integrasi Metode Resampling Dan K-Nearest Neighbor Pada Prediksi Cacat Software Aplikasi Android," *Journal Information System Development*, vol. 2, no. 1, pp. 2477–863, 2017.
- [7] M. Giegrich, R. Oomen, and C. Reisinger, "K-Nearest-Neighbor Resampling for Off-Policy Evaluation in Stochastic Control," Jun. 2023.
- [8] W. Nugraha and R. Sabaruddin, "Teknik Resampling untuk Mengatasi Ketidakseimbangan Kelas pada Klasifikasi Penyakit Diabetes Menggunakan C4.5, Random Forest, dan SVM," *Techno.COM*, vol. 20, no. 3, pp. 352–361, 2021.
- [9] M. Sari, "Penelitian Kepustakaan (Library Research) dalam Penelitian Pendidikan IPA," *NATURAL SCIENCE: Jurnal Penelitian Bidang IPA dan Pendidikan IPA*, vol. 6, no. 1, p. 41, 2020.
- [10] S. Hafni Sahir, *Metodologi Penelitian*. 2022.
- [11] P. Agustia Rahayuningsih, R. Maulana, P. Studi Komputerisasi Akuntansi, A. BSI Pontianak, and J. Abdurahman Saleh, "Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner," vol. VI, no. 1, 2018.
- [12] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [13] N. Gali, R. Mariescu-Istodor, D. Hostettler, and P. Fränti, "EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION," *Expert Systems with Applications*, vol. 129, pp. 169–185, 2019, doi: 10.1016/j.eswa.2019.03.048.
- [14] J. Mijalkovic and A. Spognardi, "Reducing the False Negative Rate in Deep Learning Based Network Intrusion Detection Systems," *Algorithms*, vol. 15, no. 8, Aug. 2022, doi: 10.3390/a15080258.
- [15] Sir, Yosua Alberth and Soepranoto, Agus H H, "Pendekatan Resampling Data Untuk Menangani Masalah Ketidakseimbangan Kelas," *Jurnal Komputer dan Informatika*, vol. 10, no. 1, pp.31-38,2022.