

PREDIKSI HUNIAN HOTEL MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS (STUDI KASUS : HOTEL RUMAH KITA KOTA CIREBON)

Hans Christian Anderson Wint's II, Ade Irma Purnama, Tati Suprapti

Teknik Informatika, STMIK IKMI Cirebon
Jalan Perjuangan No. 10 B Cirebon, Indonesia
mianacrb@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk menerapkan teknik data mining, termasuk algoritma K-Nearest Neighbors (K-NN), untuk memprediksi tingkat hunian hotel dari tahun 2021 hingga 2022 di hotel Rumah Kita yang terletak di jalan Siliwangi Kota Cirebon. Data yang digunakan dalam penelitian ini mencakup variabel seperti harga kamar, lokasi geografis, sarana hotel, dan faktor lain yang mempengaruhi tingkat hunian. Metode K-Nearest Neighbors (K-NN) digunakan untuk mengetahui tren dan hubungan antara variabel-variabel tersebut dengan tingkat hunian hotel. Data tahun 2021 dan 2022 digunakan sebagai data pelatihan untuk mengembangkan model prediksi, sementara data tahun 2023 digunakan untuk menguji akurasi model. Untuk mengukur efektivitas dan akurasi sistem, penelitian ini melakukan pengujian dengan menggunakan data aktual tahun 2023 selama periode tiga bulan terakhir. Pengujian dilakukan dengan membandingkan hasil prediksi model terhadap data sebenarnya untuk menghitung tingkat akurasi. Hasil pengujian menunjukkan bahwa model dapat memperkirakan tingkat okupansi hotel pada tahun 2023 dengan tingkat akurasi mencapai 91,67%. Prediksi hunian ramai mencapai 6 dengan akurasi class prediction 100% dan class recall 85,71%, sedangkan prediksi hunian sepi mencapai 5 dengan akurasi class prediction 83,33% dan class recall 100%. Hasil ini menunjukkan bahwa model K-NN yang dikembangkan efektif dalam memprediksi tingkat hunian dan dapat membantu manajemen hotel dalam perencanaan kapasitas, strategi penetapan harga, dan pengelolaan sumber daya. Penelitian ini berpotensi untuk diterapkan secara luas di industri perhotelan dan bisnis serupa lainnya yang mengandalkan perkiraan tingkat hunian. Dengan penambahan informasi tentang pengujian dan pengukuran akurasi, abstrak ini sekarang memberikan gambaran yang lebih lengkap tentang penelitian dan hasilnya.

Kata Kunci: Data Mining, Prediksi Hunian, Metode K-Nearest Neighbors, Industri Perhotelan.

1. PENDAHULUAN

Industri perhotelan memegang peran penting dalam perekonomian global, dengan hotel menjadi tujuan utama bagi pelancong, pembisnis, dan wisatawan yang mencari akomodasi selama perjalanan mereka. Kunci kesuksesan dalam bisnis perhotelan terletak pada pengelolaan kapasitas kamar secara optimal, antisipasi terhadap fluktuasi permintaan, dan penawaran harga yang kompetitif kepada pelanggan. Tingkat hunian hotel menjadi parameter utama yang mencerminkan efisiensi operasional dan keuntungan dalam industri ini [1].

Dalam era teknologi yang berkembang pesat ini, diharapkan dapat membantu industri perhotelan untuk lebih berkembang dan lebih siap dalam menghadapi pasang surutnya pengunjung hotel di masa depan. Perkembangan teknologi dan ketersediaan data yang melimpah dalam beberapa tahun terakhir telah memberikan kesempatan baru bagi hotel untuk meningkatkan efisiensi mereka. Data tentang reservasi kamar, harga, ulasan pelanggan, musim, dan faktor lainnya dapat diakses dengan mudah [2].

Penelitian ini bertujuan untuk mengembangkan model prediksi tingkat hunian hotel menggunakan algoritma K-Nearest Neighbor (K-NN). Algoritma K-NN adalah teknik pengambilan keputusan di mana hasil dari data masukan yang baru diklasifikasikan berdasarkan jarak terdekat dari data nilai. Algoritma

ini juga termasuk dalam algoritma supervised learning yang menghasilkan hasil dari pertanyaan instan. Hasil penelitian ini dapat memberikan kontribusi yang signifikan bagi industri perhotelan dalam mengoptimalkan operasional mereka, meningkatkan keuntungan, dan memberikan pengalaman yang lebih baik kepada pelanggan.

Dalam penelitian ini, metode pendekatan yang digunakan adalah metode pendekatan kuantitatif. Pendekatan kuantitatif menggunakan data berbasis angka dan statistik. Ini melibatkan pengumpulan data kuantitatif dan analisis statistik untuk membuat kesimpulan. Data-data tamu hotel yang digunakan berdasarkan buku arsip dan aplikasi reddorz, data-data tersebut mencakup jumlah tamu, jumlah pendapatan, dan sarana yang tersedia di hotel Rumah Kita. Data yang digunakan yaitu data dari tahun 2021 hingga tahun 2022. Dan survei pengujian metode ini dilakukan dalam 3 bulan yang di mulai dari bulan Juli tahun 2023 hingga Oktober tahun 2023 [3].

Dengan mempertimbangkan konteks dan tujuan penelitian ini, paragraf tambahan dapat berfokus pada pentingnya penelitian ini dalam konteks yang lebih luas dan bagaimana hasilnya dapat digunakan oleh industri perhotelan. Berikut adalah paragraf tambahan yang dapat Anda pertimbangkan:

Penelitian ini dapat memberikan wawasan baru dan berharga bagi industri perhotelan, khususnya

dalam konteks penggunaan teknologi dan data untuk meningkatkan efisiensi dan profitabilitas. Dengan memahami dan memprediksi tingkat hunian hotel dengan lebih akurat, hotel dapat membuat keputusan yang lebih tepat dan tepat waktu tentang pengelolaan kapasitas kamar dan penetapan harga. Selain itu, penelitian ini juga dapat membantu hotel dalam merespons dengan lebih cepat dan efektif terhadap perubahan permintaan pasar. Dengan demikian, penelitian ini tidak hanya relevan bagi Hotel Rumah Kita di Kota Cirebon, tetapi juga dapat memberikan kontribusi yang signifikan bagi industri perhotelan secara lebih luas. Melalui penelitian ini, kita dapat lebih memahami bagaimana teknologi dan data dapat digunakan untuk menghadapi tantangan dan memanfaatkan peluang dalam industri perhotelan di era digital ini.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Berdasarkan penelitian yang dilakukan oleh Septian et al., (2019), sepak bola Indonesia telah menjadi subjek banyak kontroversi dalam beberapa tahun terakhir. Ada pengurangan skor, penggantian pelatih tim nasional senior, dan penggantian Ketua Umum PSSI. Banyak pengguna jejaring sosial, termasuk Twitter, berbicara tentang masalah yang terjadi dalam beberapa bulan terakhir. Pengguna dapat membaca dan berbagi pesan singkat hingga 280 karakter di Twitter, yang merupakan platform media sosial terpopuler dalam kategori mikroblog. Berangkat dari masalah tersebut, perlu dibangun suatu sistem analisis sentimen agar dapat dengan mudah mengidentifikasi sentimen dari setiap perspektif dan pendapat dalam setiap kalimat. Analisis sentimen adalah proses menentukan apakah sebuah dokumen atau kalimat mengandung sentimen positif atau negatif. Akibatnya, akun Twitter resmi PSSI dapat digunakan sebagai sumber data untuk studi analisis emosional karena banyaknya pendapat dan pandangan pengguna Twitter tentang kontroversi yang terjadi di sepak bola Indonesia. Studi ini menggunakan data tweet bahasa Indonesia tahun 2000 yang akan dibobotkan kata menggunakan TF-IDF dan diklasifikasikan menggunakan K-Nearest Neighbors, algoritma yang cukup baik untuk mengklasifikasikan data teks [5].

Menurut penelitian yang dilakukan oleh Akbar dan Kusumodestoni (2020), tujuan utama penelitian ini adalah untuk mengoptimalkan parameter offset dan menentukan nilai k optimal untuk algoritma K-Nearest Neighbors ketika diterapkan pada data hunian tamu hotel historis. Untuk mencapai hal ini, para peneliti menyusun dataset yang terdiri dari data hunian hotel historis, yang berfungsi sebagai data pelatihan yang diawasi. Setiap baris dalam dataset diberi jumlah kolom tertentu berdasarkan parameter offset dan jumlah target yang diprediksi. Untuk menetapkan algoritma K-Nearest Neighbors, teknik validasi silang 10 kali lipat digunakan, yang melibatkan

memvariasikan nilai k dari 1 hingga 30. Melalui serangkaian percobaan, hasilnya menunjukkan bahwa pergeseran optimal dalam parameter offset dicapai dalam kisaran 14-17. Selain itu, ditemukan bahwa nilai yang paling sesuai untuk parameter k , untuk secara akurat memprediksi tingkat hunian dalam 1, 3, 6, 9, dan 12 bulan mendatang, berada dalam kisaran 5-13. Menariknya, nilai k terbaik yang diperoleh dari percobaan tidak mematuhi aturan konvensional yang berkaitan dengan akar kuadrat dari jumlah sampel data. Ini menunjukkan bahwa mungkin ada faktor lain yang mempengaruhi pemilihan nilai k optimal, yang harus dieksplorasi lebih lanjut dalam studi masa depan. Secara keseluruhan, penelitian ini menyoroti pentingnya mengoptimalkan parameter offset dan memilih nilai k yang sesuai untuk secara efektif memprediksi hunian tamu hotel berdasarkan data historis [6].

2.2. Tingkat Hunian Hotel

Tingkat hunian hotel adalah indikator penting dalam industri perhotelan yang mengukur persentase kamar yang terisi dalam periode tertentu. Tingkat hunian ini dipengaruhi oleh berbagai faktor, termasuk promosi dan periklanan, seperti yang ditunjukkan dalam penelitian terdahulu. Studi lain menunjukkan bahwa tingkat hunian hotel juga dipengaruhi oleh jumlah wisatawan dan jumlah objek wisata di suatu area [7].

2.3. Data Mining dalam Industri Perhotelan

Data mining dalam industri perhotelan merupakan proses analitis yang bertujuan untuk mengekstrak informasi berharga dan mengidentifikasi pola dari kumpulan data yang besar, yang dapat mencakup data pelanggan, transaksi penjualan, dan interaksi layanan. Dalam konteks perhotelan, data mining digunakan untuk memahami kebiasaan dan preferensi pelanggan, memprediksi tren permintaan, dan meningkatkan kualitas layanan. Sebagai contoh, penelitian yang dilakukan di Hotel Grand Zuri Dumai menunjukkan bahwa dengan menggunakan algoritma C4.5, hotel dapat menilai tingkat kepuasan pelanggan berdasarkan faktor-faktor seperti pelayanan, keramahan, kenyamanan, dan sarana yang disediakan. Hasil ini membantu hotel dalam meningkatkan layanan dan kepuasan pelanggan. Teknik data mining juga telah diterapkan dalam analisis booking hotel untuk mengevaluasi tingkat keberhasilan hotel dan dalam pengelompokan wisatawan asing ke Indonesia menggunakan algoritma K-Means, yang dapat membantu pemerintah dan industri perhotelan dalam merancang strategi pemasaran yang lebih efektif [8][9].

2.4. KNN

K-Nearest Neighbor (K-NN) adalah algoritma pembelajaran mesin yang mengklasifikasikan titik data baru dengan mempertimbangkan mayoritas kelas dari K tetangga terdekatnya. Cara kerja K-NN dimulai

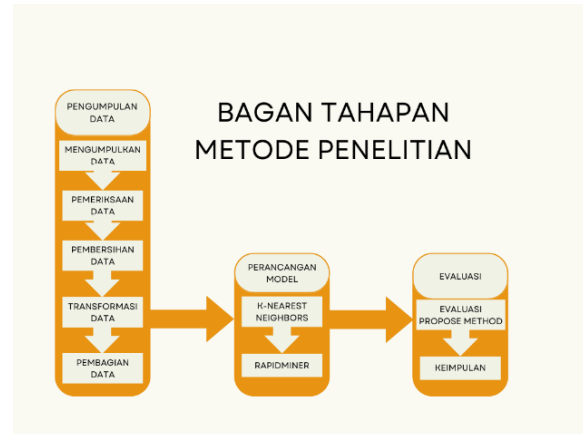
dengan memilih nilai K yang optimal, yang penting untuk menghindari sensitivitas terhadap noise atau kehilangan batasan kelas yang jelas. Setelah itu, algoritma menghitung jarak antara titik data baru dan semua titik dalam set pelatihan menggunakan metode seperti Euclidean, Manhattan, Minkowski, atau Chebyshev. Berikutnya, K tetangga terdekat diidentifikasi, dan untuk klasifikasi, titik data baru diklasifikasikan ke dalam kelas yang paling sering muncul di antara tetangga ini, sedangkan untuk regresi, diambil rata-rata dari nilai tetangga. Tahapan K-NN meliputi pemilihan nilai K , penghitungan jarak, identifikasi tetangga, voting atau penghitungan rata-rata, dan akhirnya klasifikasi atau prediksi. Penelitian menunjukkan bahwa pemilihan metode jarak dan nilai K yang tepat sangat mempengaruhi kinerja K-NN dalam berbagai aplikasi, termasuk analisis sentimen, dengan optimasi fitur dan pemilihan parameter yang cermat dapat meningkatkan akurasi dan efisiensi algoritma.[10].

2.5. Prediksi

Prediksi adalah proses mengestimasi hasil di masa depan berdasarkan data dan informasi yang ada, yang sangat penting dalam berbagai bidang seperti ekonomi, meteorologi, dan kesehatan. Proses ini melibatkan penggunaan model matematika, statistik, atau algoritma pembelajaran mesin untuk menganalisis data historis dan membuat perkiraan tentang masa depan. Prediksi kualitatif dan kuantitatif adalah dua jenis utama, dengan yang pertama berfokus pada penilaian ahli atau opini untuk hasil yang tidak dapat diukur secara numerik, seperti tren mode atau hasil pemilihan umum, dan yang kedua menggunakan data numerik untuk membuat perkiraan spesifik, seperti penjualan atau harga saham. Prediksi jangka pendek dan jangka panjang merujuk pada periode waktu prediksi, dengan jangka pendek untuk beberapa hari atau minggu dan jangka panjang untuk bulan atau tahun ke depan. Proses prediksi meliputi pengumpulan data, pemilihan dan pelatihan model, evaluasi model, dan akhirnya, menggunakan model yang telah teruji untuk prediksi masa depan. Algoritma seperti Naïve Bayes, Decision Trees, Random Forests, K-Nearest Neighbors (KNN), dan Least Squares sering digunakan, tergantung pada jenis dan tujuan prediksi. Prediksi membantu dalam pengambilan keputusan dan perencanaan strategis.[6].

3. METODE PENELITIAN

Pada metode penelitian ini akan dijelaskan Langkah-langkah yang dilakukan untuk prediksi hunian hotel yang memiliki beberapa tahapan yang di interpresentasikan dalam bentuk gambar 1.



Gambar 1. Tahapan Metode Penelitian

3.1. Pengumpulan Data

Dalam penelitian ini, pengumpulan data merupakan langkah awal yang krusial untuk membangun model prediksi tingkat hunian hotel. Data yang diperoleh berasal dari dua sumber utama:

- Buku Arsip Resepsionis: Ini adalah catatan fisik yang dijaga oleh resepsionis hotel yang mencatat detail mengenai pengunjung walk-in. Data ini mencakup jumlah pengunjung yang datang langsung ke hotel tanpa reservasi terlebih dahulu pada setiap bulan selama tahun yang diteliti.
- Aplikasi Reddorz: Merupakan platform digital yang digunakan oleh hotel untuk mencatat penjualan kamar. Data yang diambil dari aplikasi ini mencakup jumlah pengunjung yang melakukan pemesanan melalui Reddorz pada setiap bulan selama tahun yang diteliti.

Sebelum pengumpulan data, peneliti meminta izin kepada pemilik hotel untuk mengakses dan menggunakan data tersebut. Setelah mendapatkan persetujuan, peneliti memastikan bahwa data pribadi pengunjung tidak akan dimasukkan dalam dataset untuk menghindari pelanggaran undang-undang terkait pencurian data pribadi. Data yang dikumpulkan kemudian akan dianalisis untuk menentukan atribut-atribut yang relevan dengan prediksi tingkat hunian hotel. Atribut yang tidak diperlukan, seperti jumlah televisi dan jumlah AC, akan disingkirkan. Atribut yang dipertahankan dan dianalisis meliputi jumlah pengunjung walk-in dan jumlah pengunjung melalui Reddorz pada setiap bulan selama tahun yang diteliti. Informasi ini akan digunakan untuk membangun model prediksi yang akurat mengenai tingkat hunian hotel.

3.2. Transformasi

Transformasi data adalah proses mengubah data ke dalam format yang sesuai untuk analisis. Dalam penelitian ini, transformasi melibatkan normalisasi atau standarisasi nilai data untuk memastikan bahwa model K-NN dapat memproses data dengan efektif. Transformasi ini penting karena algoritma K-NN mengandalkan pengukuran jarak antar titik data, dan

perbedaan skala dapat mempengaruhi hasil prediksi. Oleh karena itu, data harus diubah menjadi format yang seragam sebelum dilakukan prediksi tingkat hunian hotel.

3.3. Perancangan Model

Perancangan model melibatkan pengembangan model prediksi menggunakan algoritma K-Nearest Neighbors (K-NN), yang merupakan metode pembelajaran mesin untuk klasifikasi dan regresi. Algoritma ini memprediksi nilai target berdasarkan kedekatan karakteristik data input dengan data latih. Dalam penelitian ini, model K-NN dirancang dengan memilih jumlah tetangga terdekat yang optimal dan menentukan metrik jarak yang sesuai. RapidMiner digunakan sebagai alat untuk merancang dan menguji model, memungkinkan otomatisasi proses dan memudahkan evaluasi kinerja model.

3.4. Evaluasi

Evaluasi model merupakan tahap penting untuk mengukur efektivitas model prediksi yang telah dikembangkan. Dalam penelitian ini, evaluasi dilakukan dengan menggunakan metrik kinerja seperti akurasi, presisi, recall, dan F1-score untuk menilai seberapa baik model K-NN dalam memprediksi

tingkat hunian Hotel Rumah Kita Kota Cirebon. Evaluasi ini membantu dalam memahami kekuatan dan kelemahan model serta dalam melakukan penyesuaian yang diperlukan untuk meningkatkan kinerjanya.

4. HASIL DAN PEMBAHASAN

Memuat hasil, Pengujian dan pembahasan tentang skripsi yang telah dilakukan

4.1. Data Selection

Proses persiapan data meliputi pengumpulan data dari sumber-sumber seperti buku arsip resepsionis dan aplikasi pemesanan kamar reddorz. Data yang dikumpulkan mencakup informasi seperti jumlah tamu, pendapatan, dan sarana hotel. Setelah pengumpulan data, dilakukan pemeriksaan dan pembersihan data untuk memastikan validitas dan kelengkapan data. Data yang tidak valid atau tidak lengkap harus disingkirkan. Selanjutnya, data mentah ditransformasikan ke dalam format yang sesuai untuk pemodelan, misalnya dengan melakukan normalisasi data agar sesuai untuk diolah oleh algoritma pemodelan. Hasil dari proses ini dapat dilihat pada tabale 1

Table 1. Data Mentah

Data Penjualan Kamar Hotel Walk-in												
Tahun	Bulan											
	Januari	Februari	Maret	April	Mei	Jun	Jul	Agustus	September	Oktober	November	Desember
2021	375	192	174	204	184	193	187	273	249	246	268	284
2022	396	186	180	228	223	219	211	254	246	220	257	258

Data Penjualan Kamar Hotel Aplikasi Reddors												
Tahun	Bulan											
	Januari	Februari	Maret	April	Mei	Jun	Jul	Agustus	September	Oktober	November	Desember
2021	421	278	248	273	213	226	218	305	277	293	302	395
2022	468	261	236	248	243	248	233	287	268	273	296	384

4.2. Pembagian Data

Data yang diubah harus dipartisi menjadi dua kumpulan data yang berbeda, yaitu data pelatihan dan data uji. Selanjutnya, atribut yang digunakan untuk data pelatihan mencakup bulan, 2021 Walk-in, 2021 Reddorz, dan keterangan. Atribut ini berfungsi sebagai fitur penting yang berkontribusi pada pelatihan model dan pada akhirnya meningkatkan kemampuan prediktifnya. Hasil Dari pembagian data latih dapat dilihat pada tabel 2

Table 2. Data Latih 2021

Bulan	Walk-in	Reddorz	Keterangan
Januari	375	421	Ramai
Februari	263	357	Ramai
Maret	174	248	Sepi

Bulan	Walk-in	Reddorz	Keterangan
April	183	256	Sepi
Mei	184	213	Sepi
Juni	193	226	Sepi
Juli	187	218	Sepi
Agustus	273	305	Ramai
September	249	277	Ramai
Oktober	246	293	Ramai
November	268	302	Ramai
Desember	284	395	Ramai

Untuk data uji atribut yang digunakan yaitu bulan, 2022 walk-in, 2022 reddorz, dan keterangan. Dapat dilihat pada Tabel 3

Table 3. Data Uji 2022

Bulan	Walk-in	Reddorz	Keterangan
Januari	396	468	Ramai
Februari	247	261	Ramai
Maret	180	236	Sepi
April	186	241	Sepi
Mei	223	243	Ramai
Juni	175	225	Sepi
Juli	211	233	Sepi
Agustus	254	287	Ramai
September	246	268	Ramai
Oktober	187	237	Sepi
November	257	296	Ramai
Desember	274	284	Ramai

Setelah data dibagi menjadi dua set yang berbeda, yaitu data pelatihan dan data uji, sekarang dalam keadaan di mana data dapat dikenakan serangkaian prosedur analitis dan teknik komputasi dengan tujuan akhir memprediksi tingkat hunian hotel secara akurat.

4.3. K-Nearest Neighbors

Dalam *k-Nearest Neighbors* penulis berfokus pada perhitungan euclidean. Euclidean sendiri merupakan proses untuk menghitung jarak antar titik dalam ruang euclidean, rumus ini menggunakan *teorema pythagoras* untuk menghitung akar kuadrat dari jumlah selisih kuadrat variable – variable antar dua titik tersebut. untuk rumus euclidean distance yang digunakan yaitu

$$d(x, y) = \sqrt{(\sum(x_i - y_i)^2) + (x_2 - y_2)^2}$$

Dimana:

- 1) $d(x, y)$ adalah jarak antara data 2021 dan data 2022
- 2) x_i adalah nilai variabel ke- i pada data 2021
- 3) y_i adalah nilai variabel ke- i pada data 2022

Menghitung jarak dengan rumus ini, akan:

- 1) Mengambil setiap pasangan nilai variabel yang sesuai dari kedua titik.
- 2) Menghitung selisih kuadrat dari pasangan nilai tersebut.
- 3) Menjumlahkan seluruh selisih kuadrat tersebut.
- 4) Mengambil akar kuadrat dari jumlah tersebut untuk mendapatkan jarak Euclidean.

Dalam konteks K-NN, jarak ini digunakan untuk menentukan K tetangga terdekat dari sebuah titik data yang tidak diketahui kelasnya (dalam kasus klasifikasi) atau nilai propertinya (dalam kasus regresi). Kemudian, berdasarkan informasi dari K tetangga terdekat ini, kelas atau nilai properti dari titik data yang tidak diketahui tersebut dapat diprediksi.

4.4. Rapid Miner

Dalam penelitian ini, model RapidMiner digunakan untuk melakukan validasi silang dan mengevaluasi akurasi prediksi tingkat hunian hotel.

Proses ini dimulai dengan langkah 'Read Excel', di mana dataset dipanggil untuk diolah. Selanjutnya, 'Set Role' digunakan untuk menetapkan peran atribut dalam dataset, yang merupakan langkah krusial untuk menghasilkan prediksi yang akurat. Langkah ketiga adalah penerapan model klasifikasi 'K-NN' yang akan digunakan dalam penelitian ini. 'Apply Model' adalah langkah keempat yang berfungsi untuk menerapkan model K-NN pada data, memungkinkan model untuk mempelajari hubungan antar atribut. Tahap terakhir, 'Performance', mengukur kinerja model, dengan output yang ditunjukkan pada gambar 2 dan 3.

	Actual	Predicted	Count
Sepi	Sepi	Sepi	236
	Ramai	Sepi	468
Ramai	Sepi	Ramai	241
	Ramai	Ramai	287

Overall Accuracy: 85.71%

Gambar 2. Hasil Peformance Rapid Miner

Metric	Value
Accuracy	91.67%
Recall	85.71%
F1 Score	88.69%

Gambar 3. Hasil Akurasi Rapid Miner

Berdasarkan gambar 3, model mencapai tingkat akurasi sebesar 91,67%, dengan prediksi hunian ramai sebanyak 6 kali dan akurasi prediksi kelas mencapai 100% serta recall kelas sebesar 85,71%. Untuk prediksi hunian sepi, terdapat 5 prediksi dengan akurasi prediksi kelas sebesar 83,33% dan recall kelas 100%.

4.5. Evaluasi

Berdasarkan hasil pengujian performa algoritma k-Nearest Neighbors pada data latih dan data uji, berikut ini akan dibahas mengenai performa algoritma serta faktor-faktor yang mempengaruhi akurasi algoritma pada kasus prediksi tingkat hunian hotel RUMAH KITA KOTA CIREBON. Performa algoritma k-Nearest Neighbors dalam mengklasifikasikan data hunian tahun 2021 dan tahun 2022 dapat dilihat dari penghitungan rumus euclidean distance dan hasil dari rapidminer. Pada kasus hunian hotel ini memiliki 2 variable yaitu bulan dengan jumlah hunian hotel. Oleh karna itu, dapat dihitung menggunakan jarak euclidean sebagai berikut:

$$\begin{aligned} \text{Januari (2021, 2022)} \\ &= \sqrt{((375 - 396)^2 + (421 - 468)^2)} \\ &= \sqrt{(441 + 2209)} = \sqrt{2650} = 51.47 \end{aligned}$$

$$\begin{aligned} \text{Februari (2021, 2022)} \\ &= \sqrt{((263 - 247)^2 + (375 - 261)^2)} \\ &= \sqrt{(256 + 13049)} = \sqrt{13305} = 51.23 \end{aligned}$$

$$\begin{aligned} \text{Maret (2021, 2022)} \\ &= \sqrt{((174 - 180)^2 + (248 - 236)^2)} \\ &= \sqrt{(36 + 144)} = \sqrt{180} = 6.32 \end{aligned}$$

$$\begin{aligned} \text{April (2021, 2022)} \\ &= \sqrt{((183 - 186)^2 + (256 - 241)^2)} \\ &= \sqrt{(9 + 225)} = \sqrt{234} = 15.29 \end{aligned}$$

$$\begin{aligned} \text{Mei (2021, 2022)} \\ &= \sqrt{((184 - 223)^2 + (213 - 243)^2)} \\ &= \sqrt{(1521 + 900)} = \sqrt{2421} = 49.20 \end{aligned}$$

$$\begin{aligned} \text{Juni (2021, 2022)} \\ &= \sqrt{((193 - 175)^2 + (226 - 225)^2)} \\ &= \sqrt{(324 + 1)} = \sqrt{325} = 18.02 \end{aligned}$$

$$\begin{aligned} \text{Juli (2021, 2022)} \\ &= \sqrt{((187 - 211)^2 + (218 - 233)^2)} \\ &= \sqrt{(576 + 225)} = \sqrt{799} = 28.30 \end{aligned}$$

$$\begin{aligned} \text{Agustus (2021, 2022)} \\ &= \sqrt{((273 - 254)^2 + (305 - 287)^2)} \\ &= \sqrt{(361 + 324)} = \sqrt{685} = 26.17 \end{aligned}$$

$$\begin{aligned} \text{September (2021, 2022)} \\ &= \sqrt{((249 - 246)^2 + (277 - 268)^2)} \\ &= \sqrt{(9 + 81)} = \sqrt{90} = 9.48 \end{aligned}$$

$$\begin{aligned} \text{Oktober (2021, 2022)} \\ &= \sqrt{((246 - 187)^2 + (293 - 237)^2)} \\ &= \sqrt{(3481 + 3136)} = \sqrt{6617} = 81.34 \end{aligned}$$

$$\begin{aligned} \text{November (2021, 2022)} \\ &= \sqrt{((268 - 257)^2 + (302 - 296)^2)} \\ &= \sqrt{(121 + 36)} = \sqrt{157} = 12.52 \end{aligned}$$

$$\begin{aligned} \text{Desember (2021, 2022)} \\ &= \sqrt{((284 - 274)^2 + (395 - 384)^2)} \end{aligned}$$

$$= \sqrt{(100 + 121)} = \sqrt{221} = 14.86$$

Dari perhitungan di atas, dapat disimpulkan bahwa jarak Euclidean antara data tahun 2021 dan tahun 2022 menunjukkan nilai yang berbeda, yang menunjukkan adanya perbedaan antara kedua tahun tersebut. Untuk mengukur akurasi prediksi, penelitian ini menggunakan model K-Nearest Neighbors dan RapidMiner. Berdasarkan hasil yang diperoleh, model mencapai tingkat akurasi sebesar 91,67%, dengan prediksi hunian ramai sebanyak 6 kali dan akurasi prediksi kelas mencapai 100% serta recall kelas sebesar 85,71%. Untuk prediksi hunian sepi, terdapat 5 prediksi dengan akurasi prediksi kelas sebesar 83,33% dan recall kelas 100%.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, algoritma K-Nearest Neighbors (K-NN) terbukti efektif dalam memprediksi tingkat hunian hotel. Model yang dihasilkan mampu memberikan perkiraan tingkat okupansi hotel pada tahun 2023 secara akurat berdasarkan data tahun sebelumnya. Hasil perkiraan ini dapat membantu manajemen hotel dalam perencanaan kapasitas, strategi penetapan harga, dan pengelolaan sumber daya. Penelitian ini juga menunjukkan kegunaan teknik data mining untuk menganalisis dan menggunakan data historis untuk peramalan dan pengambilan keputusan. Saran untuk penelitian selanjutnya meliputi penggunaan data yang lebih lengkap, mencoba algoritma machine learning lain untuk membandingkan kinerja, menerapkan model prediksi pada hotel lain, mengembangkan model prediksi untuk jangka waktu yang lebih panjang, dan mengintegrasikan model prediksi dengan sistem manajemen hotel untuk membantu pengambilan keputusan yang lebih tepat dan efektif.

DAFTAR PUSTAKA

- [1] D. I. Asmynendar, "Pengaruh Produk Domestik Regional Bruto , Tingkat Hunian Hotel , Dan Jumlah Wisatawan Terhadap Pendapatan Asli Daerah Di Kota Batu," *Al-Buhuts*, Vol. 17, Pp. 276–291, 2021.
- [2] A. Soraya, A. Pratondo, And T. D. Tambunan, "Aplikasi Pengelolaan Kamar Di Hotel Gurame Berbasis Web Application Of Room Management In Gurame Hotel Based On Web," *E-Proceeding Appl. Sci.*, Vol. 5, No. 3, Pp. 2107–2116, 2019.
- [3] E. H. Sukriadi *Et Al.*, "Pengelolaan Linen Kamar Di Internal Hotel Amaris Setiabudhi (Internal Room Linen Management Amaris Setiabudi," *Jourism Sci. J.*, Vol. 6, No. 2, Pp. 298–313, 2021, Doi: 10.32659/Tsj.V6i2.139.
- [4] J. A. Septian, T. M. Fachrudin, And A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan Tf-Idf Dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, Vol.

- 1, No. 1, Pp. 43–49, 2019, Doi: 10.52985/Insyst.V1i1.36.
- [5] J. A. Septian, T. M. Fahrudin, And A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan Tf - Idf Dan K - Nearest Neighbor,” *J. Intell. Syst. Comput.*, Pp. 43–49, 2019.
- [6] A. S. Akbar, R. H. Kusumodestoni, And R. H. Kusumodestoni, “Optimasi Nilai K Dan Parameter Lag Algoritme K-Nearest Neighbor Pada Prediksi Tingkat Hunian Hotel Optimization Of K Value And Lag Parameter Of K-Nearest Neighbor Algorithm On The,” *J. Teknol. Dan Sist. Komput.*, Vol. 8, No. January, Pp. 246–254, 2020, Doi: 10.14710/Jtsiskom.2020.13648.
- [7] D. Kepemudaan, J. Wisatawan, J. O. Wisata, And P. Sektor, “Analysis Of The Effect Of Hotel Occupancy Rates , Number Of Tourists , And Number Of Tourist Objects On The Income Of,” *Din. Dir. J. Econ.*, Vol. 1, No. 3, Pp. 294–306, 2019.
- [8] A. Budiyantera, A. K. Wijaya, And A. Gunawan, “Analisis Data Mining Hotel Booking Menggunakan Model Id3 Analysis Of Data Mining Hotel Booking Using Id3 Model,” *J. Bus. Audit Inf. Syst.*, Vol. 4, No. 1, Pp. 1–12, 2021.
- [9] T. Informatika, S. Tinggi, And T. Dumai, “Satin – Sains Dan Teknologi Informasi Penerapan Data Mining Algoritma C4 . 5 Untuk Mengetahui Tingkat Kepuasan Konsumen Di Hotel Grand Zuri Dumai,” *Satin – Sains Dan Teknol. Inf. J.*, Vol. 4, No. 2, 2018.
- [10] Z. P. Putra And A. Nugroho, “Analisa Sentimen Pengunjung Hotel Dengan K-Nearest Neighbor Studi Kasus Hotel Pop ! Surabaya,” *J. Ilmu Komput. Dan Bisnis*, Vol. 12, No. 2, Pp. 159–166, 2021.