

Perbandingan Algoritma K Nearest Neighbor Pada Nilai K untuk *Forecasting Occupancy Room* pada Hotel XYZ

Aldi Age Hasan¹, Teguh Herlambang^{2*}, M. Romli Arief³

^{1,3} Undergraduate Program Studi S1 Sistem Informasi, Fakultas Ekonomi Bisnis Teknologi Digital, Universitas Nahdlatul Ulama Surabaya

^{2*} Departemen Sistem Informasi, Fakultas Ekonomi Bisnis Teknologi Digital, Universitas Nahdlatul Ulama Surabaya

*Penulis korespondensi. Email: teguh@unusa.ac.id

ABSTRAK

Hotel memainkan peran penting sebagai faktor penunjang ekonomi sektor pariwisata di Indonesia. Menurut data Badan Pusat Statistika (BPS) Jawa Timur pada April 2023, kunjungan wisatawan mancanegara ke Jawa Timur melalui Bandara Juanda, Surabaya, meningkat 74% dibandingkan tahun sebelumnya di bulan yang sama. Tingkat hunian hotel, yang merupakan indikator jumlah kamar terisi oleh tamu, menjadi kritis dalam mengelola sumber daya seperti listrik. Penelitian ini membahas metode prediksi tingkat hunian hotel, dengan fokus pada algoritma Regresi K-NN. KNN memiliki konsistensi kuat dan efektivitas terhadap dataset pelatihan dengan banyak noise, meskipun kinerjanya dipengaruhi oleh nilai k yang menentukan sejauh mana pencarian berlangsung. Hasil yang diberikan pada penelitian ini ialah pada pembagian data training 70% dan data testing 30%, mendapatkan hasil dengan nilai K=3 dan K=4 memiliki nilai RMSE 0.204, lalu pada pembagian data 80:20 ada di nilai K=3 memiliki nilai RMSE 0.127 dan hasil pada pembagian data 90:10 pada nilai K=3 memiliki nilai RMSE 0.080.

Kata Kunci: Hotel, Okupansi, Peramalan, KNN

ABSTRACT

Hotels plays an important role as an economic support factor for the tourism in Indonesia. Based on the data of the Central Statistics Agency (BPS) of East Java in April 2023, foreign tourist visits to East Java through Juanda Airport-Surabaya increased by 74% compared to that of the previous year in the same month. The hotel occupancy rate, which is an indicator of the number of rooms occupied by guests, becomes critical in managing resources such as electricity. And, this research discusses a method of predicting hotel occupancy rates, with a focus on the K-NN Regression algorithm. KNN has strong consistency and effectiveness against training datasets with a lot of noise, although its performance is affected by the value of k which determines the extent of the search. The results of this study are, that for the division of 70% training data and 30% testing data, at the value of K = 3 and K = 4, has an RMSE value of 0.204, then that for the division of 80:20 data at the value of K = 3 has an RMSE value of 0.127, and that for the division of 90:10 data at the value of K = 3 has an RMSE value of 0.080.

Keyword: Hotel, Occupancy, Forecasting, KNN

Article info: Submitted: 6 July 2024

Accepted: 29 November 2024

How to cite this article:

Hasan, A. A., Herlambang, T., & Arief, M. R. (2024). Perbandingan Algoritma K Nearest Neighbor Pada Nilai K untuk Forecasting Occupancy Room pada Hotel XYZ. Zeta - Math Journal, 9(2), 97-104. <https://doi.org/10.31102/zeta.2024.9.2.97-104>



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](#).

1. PENDAHULUAN

Hotel merupakan salah satu faktor penunjang pada perekonomian di Indonesia pada sektor pariwisata. Menurut sumber data BPS (Badan Pusat Statistika) Jawa Timur pada bulan April 2023, kunjungan wisatawan mancanegara ke Jawa Timur melalui pintu masuk Bandara Juanda di Surabaya mengalami peningkatan. Jumlah kunjungan ini naik 74% dari tahun sebelumnya di bulan yang sama. Sudah sepantasnya bagi pemilik hotel untuk menentukan perencanaan yang baik kedepannya agar dapat menampung jumlah wisatawan. Terlebih lagi di kota Surabaya yang merupakan kota terbesar yang ada di Jawa Timur dengan melakukan manajerial kualitas pelayanan yang baik dan juga ketersediaan jumlah kamar. Dengan meningkatkan kualitas manajerial pelayanan yang baik, maka para wisatawan akan merasa lebih aman, nyaman, dan betah untuk tinggal lebih lama di dalam hotel sehingga dapat menaikkan ekonomi didaerah Surabaya di bidang pariwisata (T. Herlambang *et al.*, 2022).

Forecasting adalah bagian awal dari proses aktivitas pengambilan suatu keputusan yang dibutuhkan untuk menentukan kapan terjadinya besarnya nilai data pada suatu peristiwa (Chang Y. M *et al.*, 2022). *Forecasting* merupakan suatu kegiatan memperkirakan atau memprediksikan kejadian dimasa yang akan datang tentunya dengan bantuan penyusunan rencana terlebih dahulu, dimana rencana ini dibuat berdasarkan kapasitas dan kemampuan permintaan/keuntungan yang telah dilakukan di perusahaan (Anshori M. Y., 2023). Adapun 3 prosedur yang penting kita ketahui dalam melakukan peramalan yakni dengan menganalisa data masa lalu untuk mengetahui pola data masa lalu sebagai perbandingan di masa depan. Lalu selanjutnya ialah menentukan data yang akan digunakan. Selanjutnya yang terakhir adalah memproyeksikan data yang lalu dengan metode yang digunakan.

Beberapa metode algoritma yang sering digunakan dalam melakukan *forecasting* yakni Regresi Linear, C-45, Naïve Bayes, SVM, Neural Network dan masih banyak lainnya. Dikutip dari hasil penelitian SLR (Studi Literatur Review) di tahun 2011 hingga 2021 yang dilakukan oleh Zhang pada tahun 2022 (W. Zhang, *et al.*, 2022), bahwa algoritma K-Nearest Neighbor (KNN) ini merupakan algoritma *machine learning* yang banyak dipakai untuk dilakukan penelitian. Pada penelitian ini akan melakukan *forecasting* pada tingkat hunian hotel dengan menggunakan algoritma dan juga K-Nearest Neighbor (KNN). Algoritma K-Nearest Neighbor (KNN) adalah algoritma yang memiliki konsistensi yang kuat dan memiliki kelebihan dalam menangani dataset pelatihan yang memiliki banyak *noise* (Labolo A.Y., *et al.*, 2023). Kelebihan lainnya dari K-Nearest Neighbor (KNN) ini adalah algoritma ini sangat efektif terhadap jumlah data pelatihan yang tinggi atau besar (Putrasyah A. *et al.*, 2023).

Pada penelitian sebelumnya, algoritma KNN ini juga sudah digunakan secara luas. Beberapa contoh penelitian yang mengimplementasikan algoritma K-Nearest Neighbor (KNN) diantaranya, algoritma K-Nearest Neighbor (KNN) berhasil mencapai tingkat akurasi 91,67% dalam memprediksi tingkat hunian hotel di kota Cirebon (A.S Akbar. *et al.*, 2020). Kemudian, ditahun sebelumnya yaitu pada tahun 2023, algoritma K-Nearest Neighbor (KNN) digunakan untuk memprediksi harga rata-rata kamar di Hotel S (Asy'ari V, *et al.*, 2023). Disamping itu, algoritma K-Nearest Neighbor (KNN) juga dapat diimplementasikan pada bidang lain seperti peramalan harga saham bank swasta (Muallif I.S. *et al.*, 2023).

Berikut merupakan tabel dataset histori okupansi pada Hotel XYZ yang memiliki rentang tahun 2018 sampai 2023.

Tabel 1. Data Hotel XYZ

Bulan	No. Of Rooms	Rooms Avail.	% Occ	Rooms Sold	Av Rm Rate	Room Revenue
1 Apr 2018	126	3780	34%	1283	370.076	474.807.004
1 May 2018	126	3780	42%	1572	334.352	541.321.824
1 Jun 2018	126	3780	20%	767	328.872	252.244.964
1 Jul 2018	126	3906	76%	2988	548.454	2.934.778.785
1 Jan 2019	153	4743	34%	1628	327.060	532.453.680
1 Feb 2019	153	4285	48%	2049	333.930	684.222.570

1 Mar 2019	153	4743	21%	2491	332.844	829.114.404
1 Apr 2019	153	4590	52%	2367	318.714	754.396.038
1 Jan 2020	153	4743	38%	1799	320.591	576.743.851
1 Feb 2020	153	4437	53%	2343	316.913	742.528.232
1 Mar 2020	153	4743	40%	1882	300.502	565.544.814
1 Apr 2020	153	4590	17%	787	222.328	174.972.466
1 Jan 2021	153	4743	330%	1566	269.479	422.004.146
1 Feb 2021	153	4284	35%	1515	267.360	405.050.299
1 Mar 2021	153	4743	32%	1524	272.150	414.757.018
1 Apr 2021	153	4590	38%	1747	276.652	483.310.559
1 Jan 2022	153	4743	55%	2625	278.432	730.882.793
1 Feb 2022	153	4284	58%	2477	262.718	650.751.857
1 Mar 2022	153	4743	67%	3199	264.537	846.253.531
...	...	...	...	...	...	...
1 Apr 2023	153	4590	69%	3178	308.834	981.475.958

2. METODE PENELITIAN

Secara umum, pada penelitian ini menggunakan metode kuantitatif dengan dataset yang didapatkan dari hasil histori Hotel XYZ yang kemudian diolah menggunakan Rapidminer dengan algoritma K-Nearest Neighbor (KNN) untuk memprediksi hasil produksi kelapa pada masa mendatang

3. HASIL PENELITIAN

3.1 Algoritma K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) adalah metode yang menggunakan algoritma *supervised* dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari *label class* pada KNN (Drajana I.C.R, *et.al.*,2023). Prinsip K-Nearest Neighbor (KNN) melibatkan pengklasifikasian setiap sampel yang mirip dengan sampel-sampel disekitarnya (M.N Muttaqin, *et.al.*,2021). Algoritma K-Nearest Neighbor (KNN) menggunakan klasifikasi ketetanggaan sebagai nilai prediksi yang dihitung menggunakan fungsi *Euclidean*. Berikut merupakan fungsi matematika dari aturan *Euclidean*.

$$D = \sqrt{(x_2 - x_1)^2 - (y_2 - y_1)^2} \quad (1)$$

dengan:

D : Jarak

x : Sample data

y : Data uji

3.2 Root Mean Squared Error (RMSE)

Metode *Root Mean Squared Error* (RMSE) merupakan metode kuadratik untuk mengukur skor dengan cara menghitung rata-rata magnitudo dari *error*. RMSE mengindikasikan beberapa hal yaitu kesesuaian absolut antara model dan data. Selain itu, RMSE menunjukkan rata-rata *error* prediksi dari model. Penelitian ini memanfaatkan metode RMSE untuk mengukur keakuratan model K-Nearest Neighbor (KNN) yang terbentuk untuk memprediksi tingkat hunian pada hotel. Semakin rendah nilai dari RMSE menunjukkan model yang dibuat mampu melakukan prediksi dengan baik (Choirun A., *et.al.*,2020). Perhitungan RMSE dapat dilakukan dengan persamaan berikut.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(X_i - F_i)^2}{n}} \quad (2)$$

dengan:

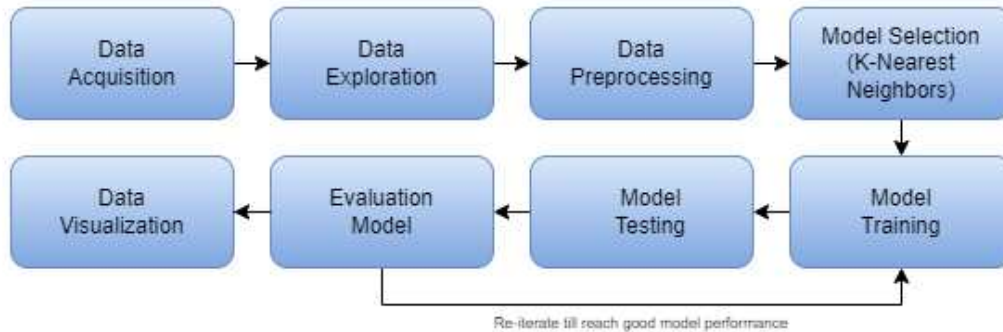
n : Banyak data hasil ramalan

X_i : Data ramalan dihitung dari model yang digunakan pada waktu / tahun i

F_i : Data sebenarnya

3.3 Implementasi K Nearest Neighbor (KNN)

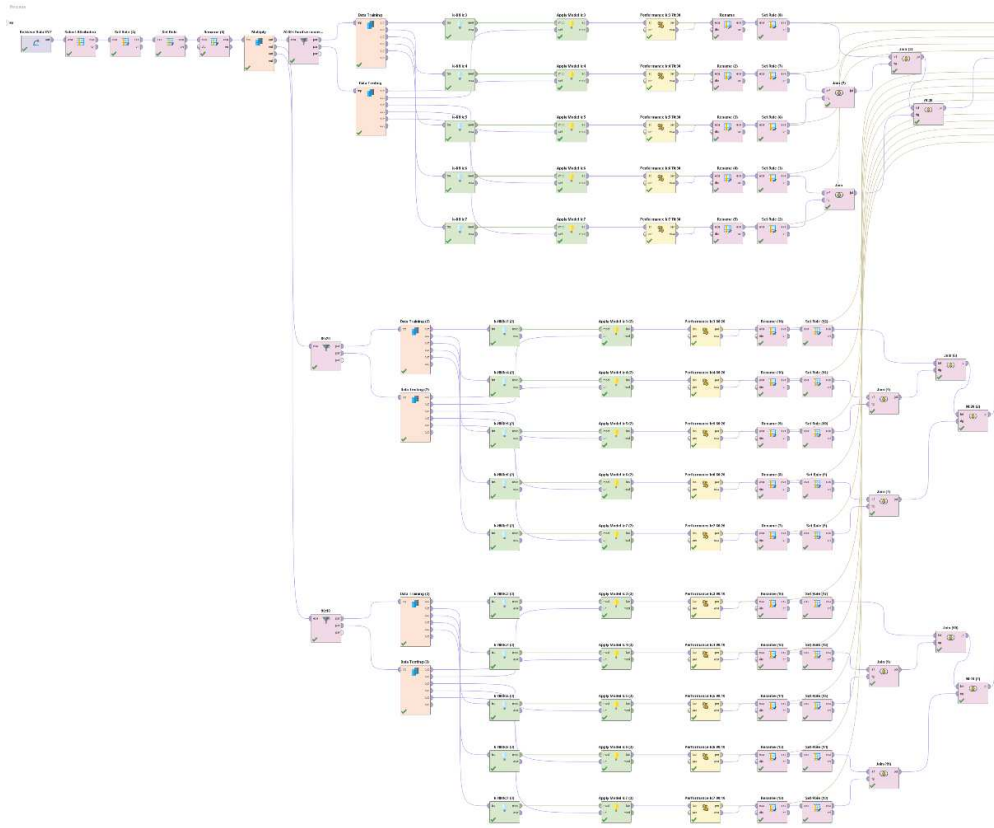
Berikut merupakan implementasi dari metode K Nearest Neighbor (KNN) pada data histori okupansi di Hotel XYZ.



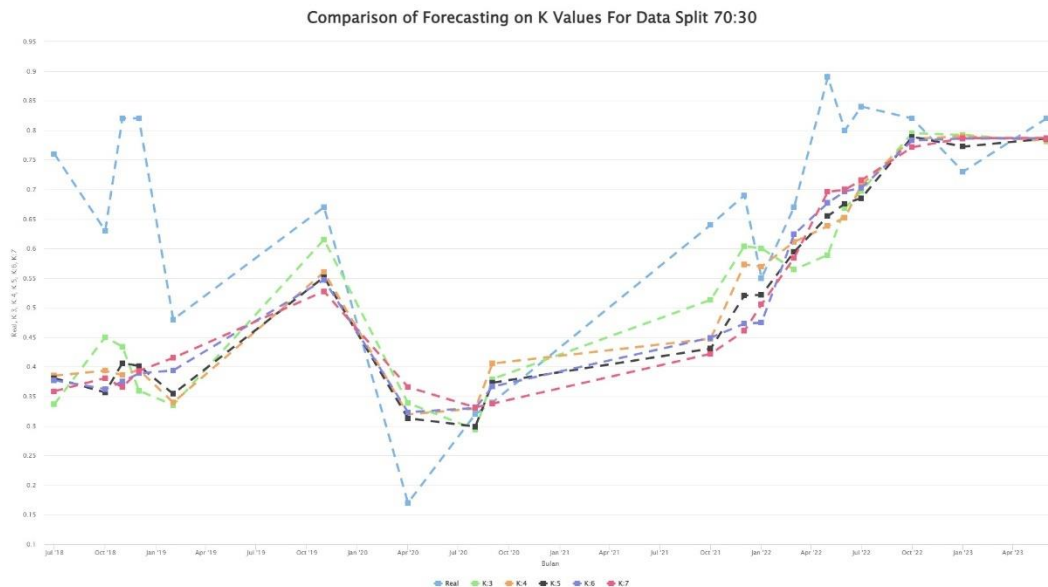
Gambar 2. Implementasi Algoritma K Nearest Neighbor (KNN)

Analisis dan prediksi dengan metode K-Nearest Neighbor (KNN) diawali dengan menentukan variabel-variabel independen dan dependen. Sebelum dilakukan simulasi dan prediksi untuk menemukan nilai keakurasiannya.

Pada penelitian ini, simulasi diawali dengan membagi dataset dari fitur/kolom-kolom yang terpilih sebagai variabel independen dan dependen kedalam *data training* dan *data testing* berdasarkan proporsi tertentu. Simulasi pertama membagi *data training* sebesar 70% dan *data testing* sebesar 30%. Simulasi kedua membagi *data training* sebesar 80% dan *data testing* sebesar 20%. Simulasi ketiga membagi *data training* sebesar 90% dan *data testing* sebesar 10%.

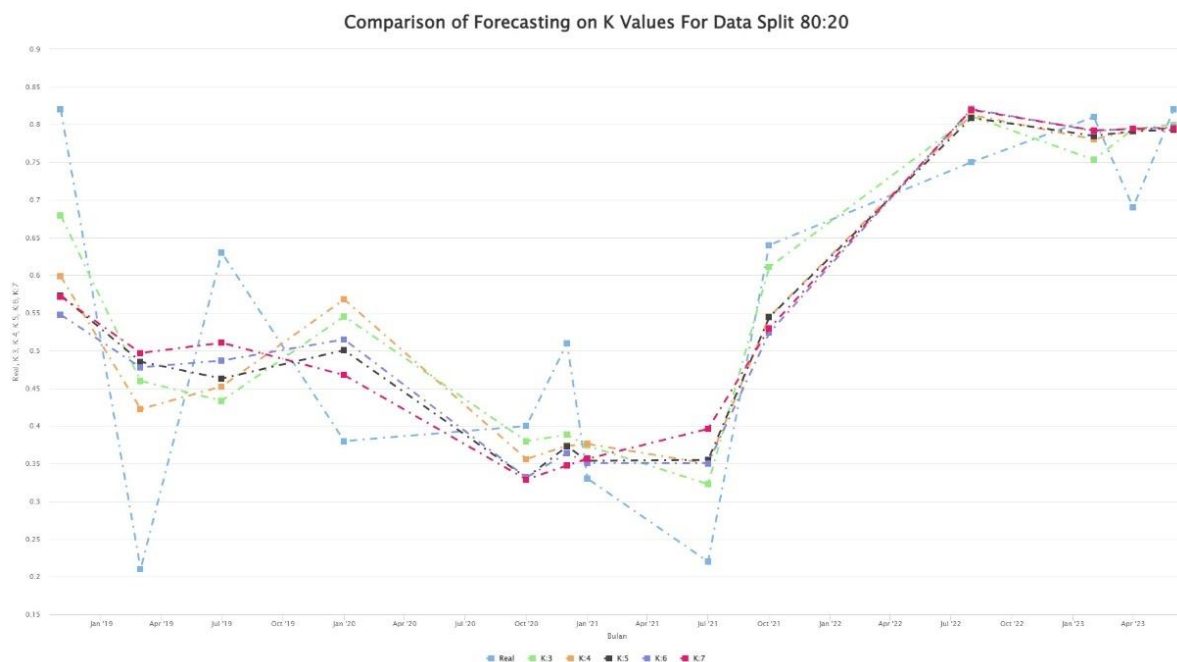


Gambar 3. Pemodelan *Forecasting* K Nearest Neighbor (KNN) Menggunakan Rapidminer



Gambar 4. Hasil Perbandingan Forecasting Pada Nilai K Dengan Data Latih 70% dan Data Uji 30%

Pada **Gambar 4** menunjukkan grafik perbandingan akurasi nilai K pada algoritma K-Nearest Neighbor (KNN) dengan pembagian data latih sebesar 70% dan data uji sebesar 30%. Grafik pada Gambar 4 menunjukkan bahwa K=3 yang digambarkan dengan garis hijau terlihat ada fluktuasi yang signifikan dalam prediksi, terutama pada periode awal pada analisis pada rentang tahun 2018 sampai tahun 2019. Selanjutnya Nilai K=4 yang digambarkan pada garis warna oranye ini menunjukkan performa yang lebih baik dibandingkan K=3, dengan prediksi yang lebih mendekati data aktual pada beberapa periode. Nilai K=5 menunjukkan keseimbangan yang baik antara stabilitas dan akurasi. Prediksi K=5 lebih dekat dengan data aktual dibandingkan K=3 dan K=4. Nilai K=6 menunjukkan performa yang sangat baik, dengan prediksi yang lebih stabil dan mendekati data aktual pada banyak titik. Nilai K=7 menunjukkan performa prediksi yang paling stabil dan akurat dibandingkan dengan nilai K lainnya.

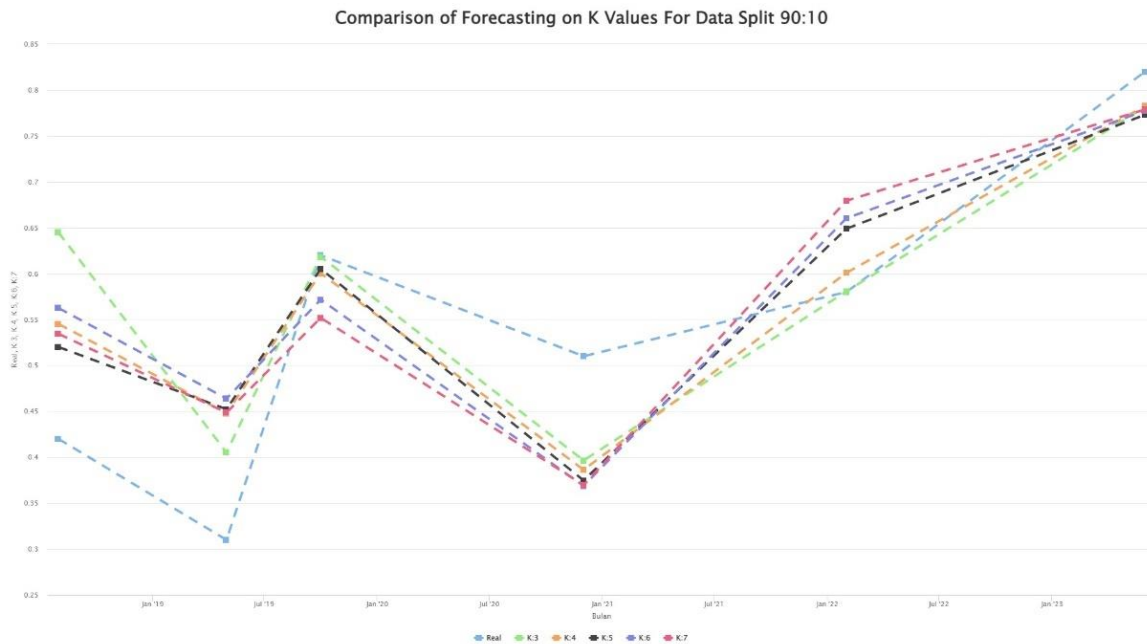


Gambar 5. Hasil Perbandingan Forecasting Pada Nilai K Dengan Data Latih 80% dan Data Uji 20%

Pada **Gambar 5** menunjukkan grafik perbandingan akurasi nilai K pada algoritma K-Nearest Neighbors (KNN) dengan pembagian data 80:20 menunjukkan bahwa Prediksi dengan nilai K=3 dengan digambarkan garis hijau, cenderung memiliki pola data fluktuatif dan tidak stabil. Nilai K=4 yang digambarkan dengan garis

oranye menunjukkan kinerja yang lebih baik dibandingkan $K=3$ dengan prediksi yang mendekati data sebenarnya.

Pada nilai $K=5$ diwakili oleh garis hitam, menunjukkan keseimbangan yang baik antara stabilitas dan akurasi, dengan prediksi lebih dekat dengan data aktual dibandingkan $K=3$ dan $K=4$. Pada nilai $K=6$ yang digambarkan dengan garis berwarna merah muda menunjukkan performa yang sangat baik dengan prediksi yang lebih stabil dan mendekati data aktual pada banyak titik. Nilai $K=7$ yang digambarkan dengan warna merah, menunjukkan performa prediksi yang paling stabil dan akurat dibandingkan nilai K lainnya dengan prediksi yang mendekati data aktual pada banyak periode.



Gambar 6. Hasil Perbandingan Forecasting Pada Nilai K Dengan Data Latih 90% dan Data Uji 10%

Pada **Gambar 6** menunjukan grafik perbandingan akurasi nilai K pada algoritma K-Nearest Neighbors (KNN) dengan pembagian data 90:10 menunjukkan bahwa Prediksi dengan nilai $K=3$ dengan digambarkan garis hijau cenderung memiliki pola data *forecasting* fluktuatif dan menunjukkan beberapa penyimpangan dari data real. Pada nilai $K=4$ digambarkan pada garis berwarna oranye ini menunjukkan performa yang lebih baik dibandingkan $K=3$, dengan nilai data prediksi yang mendekati data aktual pada beberapa periode.

Pada nilai $K=5$ yang digambarkan dengan garis berwarna hitam, yang menunjukkan keseimbangan yang baik antara stabilitas dan akurasi, dengan pola data prediksi lebih dekat dengan data aktual dibandingkan dengan nilai $K=3$ dan $K=4$. Pada nilai $K=6$ yang digambarkan dengan garis berwarna ungu, menunjukkan performa yang sangat baik dengan pola data prediksi yang lebih stabil dan mendekati data aktual pada banyak titik. Pada nilai $K=7$ yang digambarkan dengan garis berwarna merah menunjukkan performa prediksi yang memiliki pola data prediksi yang berbeda dengan data real. Jauh lebih baik dengan nilai $K=3$, $K=4$ dan $K=5$.

Setelah melakukan simulasi *forecasting* model berdasarkan pembagian data, pada Tabel 2 berikut merupakan hasil rekapitulasi dari hasil nilai kesalahan (RMSE).

Tabel 2. Tabel Rekapitulasi Nilai RMSE

Nilai K	RMSE		
	70:30	80:20	90:10
K=3	0.205	0.127	0.080
K=4	0.204	0.132	0.093
K=5	0.204	0.139	0.096
K=6	0.208	0.141	0.111
K=7	0.212	0.140	0.107

4. KESIMPULAN

Berdasarkan analisis RMSE (*Root Mean Square Error*) untuk berbagai nilai K pada algoritma K Nearest Neighbor (KNN) dengan pembagian data yang berbeda yaitu, 70%:30%, 80%:20%, dan 90%:10%, terlihat bahwa nilai K terbaik bervariasi menurut divisi data. Pada pembagian data 70%:30%, nilai RMSE terendah dicapai oleh K=4 dan K=5 dengan nilai yang sama yaitu 0,204 yang menunjukkan bahwa kedua nilai K ini memberikan prediksi yang paling akurat untuk pembagian data ini. Untuk pembagian data 80%:20% dan 90%:10%, K=3 memberikan RMSE terendah masing-masing sebesar 0,127 dan 0,080 yang menunjukkan bahwa K=3 lebih unggul pada kondisi dengan data latih yang lebih banyak, mampu menangkap pola data dengan baik dan memberikan prediksi terbaik. tepat. Nilai K yang lebih rendah seperti K=3 cenderung lebih responsif terhadap pola pada data ketika jumlah data pelatihan lebih besar, sedangkan nilai K yang lebih tinggi memberikan keseimbangan yang lebih baik antara bias dan varians pada kondisi dengan data pelatihan yang lebih sedikit.

DAFTAR PUSTAKA

- A. S. Akbar and R. H. Kusumodestoni, "Optimization of k value and lag parameter of k-nearest neighbor algorithm on the prediction of hotel occupancy rates," *J. Teknol. dan Sist. Komput.*, vol. 8, no. 3, pp. 246–254, Jul. 2020. <https://doi.org/10.14710/jtsiskom.2020.13648>
- Anshori, M. Y., Asy'Ari, V., Herlambang, T., & Farid, I. W. (2023, November). Forecasting occupancy rate using neural network at Hotel R. In *2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)* (pp. 347-351). IEEE. W. Zhang, Y. Wu, and J. K. Calautit, "A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment," *Renew. Sustain. Energy Rev.*, vol. 167, no. January, p. 112704, 2022. <https://doi.org/10.1016/j.rser.2022.112704>
- Asy'ari, V., Anshori, M. Y., Herlambang, T., Farid, I. W., Karya, D. F., & Adinugroho, M. (2023, November). Forecasting average room rate using k-nearest neighbor at Hotel S. In *2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)* (pp. 496-500). IEEE. <https://doi.org/10.1109/ICAMIMIA60881.2023.10427942>
- Chang, Y. M., Chen, C. H., Lai, J. P., Lin, Y. L., & Pai, P. F. (2021). Forecasting hotel room occupancy using long short-term memory networks with sentiment analysis and scores of customer online reviews. *Applied Sciences*, 11(21), 10291. <https://doi.org/10.3390/app112110291>
- CHOIRUN, A., & Andri, A. (2020). *PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK PREDIKSI PENJUALAN OBAT PADA APOTEK KIMIA FARMA ATMO PALEMBANG* (Doctoral dissertation, Universitas Bina Darma). I. Wayan and P. Pratama, "JAVOK: Jurnal Akademisi Vokasi IMPLEMENTASI SUPPORT VECTOR MACHINE DALAM PREDIKSI HARGA RUMAH," vol. 2, no. 2, pp. 101–113, 2023.
- Drajana, I. C. R., & Botutihe, M. H. (2023). Perbandingan Metode Forecasting K-NN, NN dan SVM Untuk Peramalan Jumlah Produksi Coconut Oil. *JURNAL TECNOSCIENZA*, 7(2), 298-311. <https://doi.org/10.51158/tecnoscienza.v7i2.919>
- Labolo, A. Y., Utiahman, S. A., Lasulika, M. E., Drajana, I. C. R., & Bode, A. (2023). K-Nearest Neighbor for Gorontalo City Chili Price Prediction Using Feature Selection, Backward Elimination, and Forward Selection. *International Journal Software Engineering and Computer Science (IJSECS)*, 3(3), 261-269., <https://doi.org/10.35870/ijsecs.v3i3.1709>
- M. N. Muttaqin and I. Kharisudin, "Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor," *UNNES J. Math.*, vol. 10, no. 2, pp. 22–27, 2021, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>
- Muallif, I. S., Budiman, H., & Ransi, N. (2023, November). Penerapan Data Mining untuk Prediksi Pergerakan Harga Saham Menggunakan Algoritma K-Nearest Neighbor. In *Prosiding Seminar Nasional*

Pemanfaatan Sains Dan Teknologi Informasi (Vol. 1, No. 1, pp. 297-306). Retrieved from <https://epublikasi.digitallinnovation.com/index.php/semptatin/article/view/46>

- Putrasyah, A., Ermatita, E., & Abdiansah, A. (2023). Klasifikasi Kemiripan Suara Rekaman Menggunakan Metode Me-Frequency Cepstra Coefficient dan Minkowski. *INTECOMS: Journal of Information Technology and Computer Science*, 6(2), 959-969.
- Purnamasari, A. I., & Suprati, T. (2024). PREDIKSI HUNIAN HOTEL MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1897-1903. <https://doi.org/10.31539/intecom.v6i2.7023>
- Putri Amaningsih, Tony Yulianto, Faisol, & Rica Amalia. (2024). Implementation of K-Means and Single Linkage on Types of Disabilities in East Java Province. *Contemporary Mathematics and Applications (ConMathA)*, 6(2), 101–113. <https://doi.org/10.20473/conmatha.v6i2.58513>
- T. Herlambang, M. Y. Anshori, V. Asy'ari, and B. P. Tomasouw, "TITLE OF ARTICLE," *together j. Math. & App*, 2022, doi: 10.30598/barekengxxxxxxxxxxxxxx.
- W. Zhang, Y. Wu, and J. K. Calautit, "A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment," *Renew. Sustain. Energy Rev.*, vol. 167, no. January, p. 112704, 2022. <https://doi.org/10.1016/j.rser.2022.112704>