

BAB II

TINJAUAN PUSTAKA

2.1. Landasan Teori

2.1.1. Prediksi

Prediksi merupakan proses untuk memperkecil kesalahan, sehingga dapat meminimalisir antara kejadian yang sebenarnya. Prediksi dapat menentukan kebutuhan dalam kurun waktu minggu, bulan, bahkan tahun hal ini berkaitan dengan dukungan jumlah data atau prediksi waktu yang dianalisa sehingga dapat diperhitungkan dalam memprediksi data pada tahun mendatang. (Faisal dkk., 2023)

2.1.2. Algoritma

Algoritma adalah serangkaian prosedur yang saling berinteraksi untuk mencapai suatu tujuan dalam memecahkan masalah tertentu. Sehingga algoritma dapat digunakan untuk penghitungan, pemrosesan data, pencarian, penalaran, optimasi, pembelajaran dan sejenisnya untuk menyelesaikan suatu masalah(Santoso dkk., 2020).

2.1.3. Preprocessing data

Pra-pemrosesan data adalah tahap penting dalam machine learning yang bertujuan untuk mempersiapkan data agar dapat digunakan oleh algoritma. Tahap ini mencakup:

- a. *Label Encoding* adalah sebuah proses mengubah data kategorikal menjadi bentuk numerik (Herdian dkk., 2024) . Contohnya, jenis sapi seperti "Madura", "Bali", dan "Simental" diubah menjadi nilai 0.2, 0.4, dan 0.6.

- b. Normalisasi adalah sebuah proses mengubah skala data numerik ke rentang tertentu (misalnya 0 sampai 1) agar fitur memiliki bobot yang seimbang (Fuji Astari dkk., 2024). Untuk rumus *MinMax* ditunjukkan pada rumus 2.1.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2.1)$$

Keterangan :

x : nilai asli

x_{min} : nilai minimum dalam data

x_{max} : nilai maksimum dalam data

x' : nilai setelah dinormalisasi (rentangnya menjadi 0 sampai 1)

2.1.4. Machine Learning

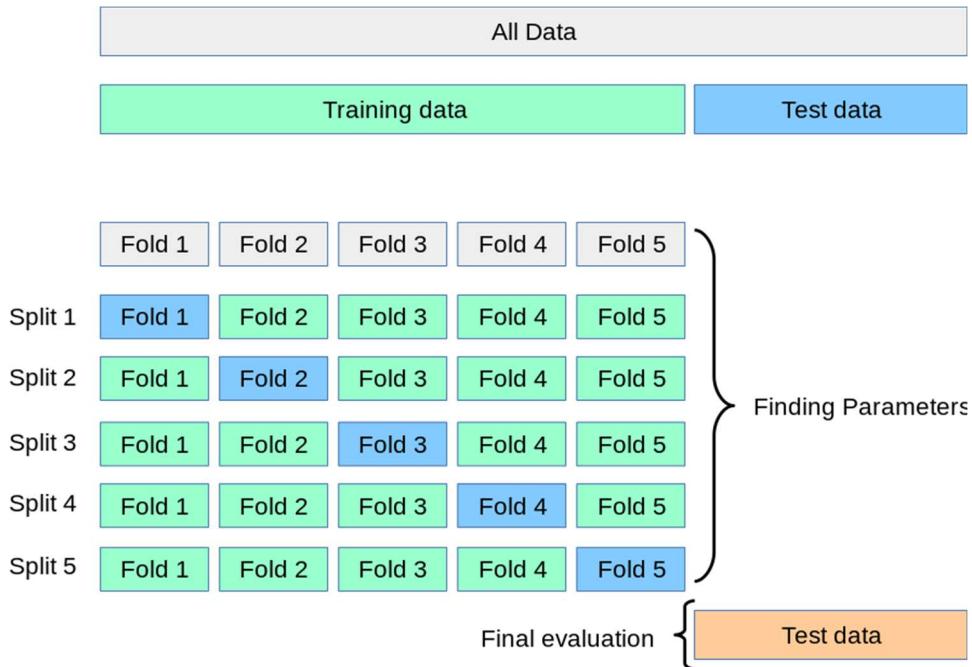
Machine Learning merupakan model statistik atau algoritma yang mampu menyelesaikan tugas tertentu tanpa perlu instruksi langsung, dengan mengandalkan pola khusus(Puspandoyo dkk., 2022). *Machine learning* adalah sistem yang dapat belajar membuat keputusan sendiri tanpa harus diprogram berulang kali oleh manusia sehingga komputer dapat menjadi lebih pintar dan belajar dari pengalamannya dengan data (Wijoyo, dkk 2024). Machine learning merupakan bagian dari kecerdasan buatan. *Machine learning* adalah teknik untuk melakukan inferensi terhadap data dengan pendekatan matematis (Diantika,Firmanto. 2020).

2.1.5. Cross Validation

Cross validation merupakan teknik resampling yang penting dalam statistik modern, yang digunakan untuk mengevaluasi kinerja model prediktif (Muhammad Iqbal, 2021). Teknik ini bekerja dengan membagi dataset menjadi dua bagian utama, yaitu data latih dan data uji (Wijiyanto dkk., 2024). Dengan cara ini, *cross validation* mampu membantu dalam menentukan parameter model yang paling optimal serta memberikan estimasi yang lebih akurat terhadap performa model dalam memprediksi data baru.

2.1.6. K-Fold Cross Validation

K-Fold Cross Validation adalah salah satu teknik validasi silang yang dirancang untuk mengatasi keterbatasan metode *holdout*, di mana dataset biasanya dibagi secara acak menjadi data latih dan data uji satu kali saja (Hafid, 2023). dalam metode *K-Fold*, dataset dibagi secara merata menjadi k *subset* atau *fold*. Proses pelatihan dan pengujian model kemudian dilakukan sebanyak k kali (iterasi), di mana pada setiap iterasi satu subset digunakan sebagai data uji (test set), sementara $k-1$ subset sisanya digunakan sebagai data latih (training set) (Mariyani, Yulita Molliq Rangkuti, 2024). dengan pendekatan ini, setiap data dalam *dataset* memiliki peluang yang sama untuk digunakan sebagai data latih maupun data uji, yang pada akhirnya menghasilkan evaluasi model yang lebih stabil dan representatif. Teknik ini juga membantu mengurangi bias yang dapat muncul akibat pembagian data yang tidak merata pada metode *holdout*. Hasil akhir dari *K-Fold Cross Validation* biasanya diperoleh dari rata-rata akurasi atau metrik evaluasi lainnya dari seluruh iterasi (Hafid, 2023).



Gambar 2. 1 *K-fold cross validation* (Jamaleddyn dkk., 2023)

Gambar 2.1 menunjukkan proses *Cross Validation*, di mana data pelatihan dibagi menjadi 5 bagian (*fold*). Setiap *fold* secara bergantian digunakan sebagai data validasi, sementara *fold* lainnya digunakan untuk melatih model. Proses ini diulang sebanyak jumlah *fold* (5 kali), lalu hasil evaluasi dirata-rata untuk memperoleh performa model yang lebih stabil. Setelah validasi selesai, model akhir diuji menggunakan data uji terpisah.

2.1.7. GridSearchCV

GridSearchCV merupakan teknik yang digunakan untuk mengidentifikasi kombinasi *hyperparameter* yang paling optimal guna meningkatkan kinerja suatu model (Nurul, 2024). Prosedur ini mengevaluasi berbagai kombinasi *hyperparameter* dengan melakukan validasi silang, sehingga dapat diketahui rata-rata performa model pada setiap konfigurasi. Melalui pendekatan sistematis ini, pengguna dapat menentukan parameter yang paling sesuai untuk diterapkan.

GridSearchCV sendiri merupakan bagian dari pustaka *scikit-learn* yang memfasilitasi proses pencarian hyperparameter terbaik dengan mengevaluasi banyak model sekaligus (Maisat et al, 2025).

2.1.8. K-Nearest Neighbors (KNN)

K-Nearest Neighbor (KNN) adalah salah satu algoritma yang sering digunakan dalam klasifikasi data mining. Algoritma ini berfungsi untuk menganalisis data dan membantu dalam proses pencarian atau prediksi terhadap hasil yang mungkin terjadi di masa mendatang, berdasarkan pola atau hubungan yang ditemukan dari data sebelumnya (Faisal dkk., 2023). Penelitian (Asrory dkk., 2020) Menerangkan bahwa Klasifikasi ketetanggaan dalam metode KNN adalah klasifikasi yang di dasari oleh jarak data baru dengan tingkat kemiripan data baru tersebut terhadap data pola.

Adapun langkah-langkah untuk menghitung Metode *k-Nearest Neighbor* adalah sebagai berikut:

1. Menentukan parameter *k*.
2. Menghitung jarak antara yang akan dievaluasi dengan semua pelatihan.
3. Mengurutkan jarak yang terbentuk.
4. Menentukan jarak terdekat sampai urutan *k*.
5. Memasangkan kelas yang bersesuaian.
6. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi.

Perhitungan jarak antara data baru dengan data lama dilakukan menggunakan rumus sebagai berikut:

$$d = \sum \sum_{i=1}^p (x_{2i} + x_{1i})^2 \quad (2.2)$$

Keterangan:

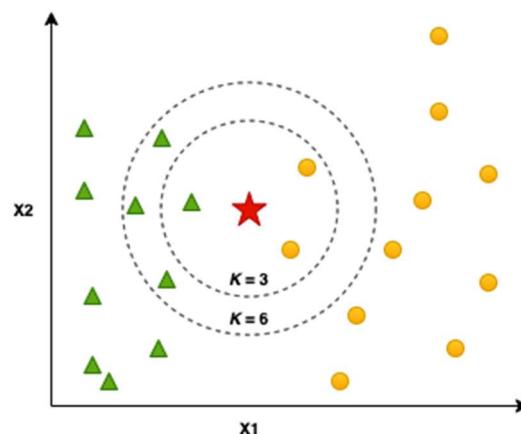
x_1 = Sampel data

x_2 = Data uji atau data testing

i = variabel data

d = jarak

p = Dimensi data



Gambar 2. 2 ilustrasi algoritma K-Nearest Neighbors (Rina, 2023)

Gambar 2.2 menunjukkan ilustrasi algoritma *K-Nearest Neighbors* (KNN) dalam menentukan kelas suatu data uji yang ditandai dengan simbol bintang merah. Data latih terdiri dari dua kelas, yaitu segitiga hijau dan lingkaran kuning. dengan $K = 3$, tiga tetangga terdekat dari titik uji terdiri dari dua segitiga dan satu lingkaran, sehingga titik tersebut diklasifikasikan ke dalam kelas segitiga hijau. Namun, saat $K = 6$, terdapat lebih banyak lingkaran (empat) dibanding segitiga (dua), sehingga klasifikasinya berubah menjadi kelas lingkaran kuning. Ilustrasi ini

memperlihatkan bagaimana pemilihan nilai K dapat memengaruhi hasil klasifikasi pada algoritma KNN.

2.1.6 Evaluasi Kinerja Model

Dalam mengevaluasi performa model prediksi regresi, digunakan beberapa metrik yang bertujuan untuk mengukur tingkat kesalahan antara hasil prediksi dan nilai actual (Hakim & Afriliansyah, 2024). Metrik-metrik tersebut meliputi *Mean Absolute Error (MAE)*, *Root Mean Square Error (RMSE)*, *Mean Absolute Percentage Error (MAPE)*, dan Koefisien Determinasi (R^2 Score) (Hakim & Afriliansyah, 2024). Masing-masing metrik memberikan sudut pandang berbeda terhadap kinerja model.

a. *Mean Absolute Error (MAE)*

MAE adalah ukuran rata-rata dari selisih absolut antara nilai prediksi dan nilai actual (Suryanto, 2019). Metrik ini menunjukkan seberapa besar kesalahan prediksi model secara umum, tanpa mempertimbangkan arah kesalahan (positif atau negatif). Semakin kecil nilai MAE, maka semakin baik performa model.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \quad (2.3)$$

Keterangan:

n = jumlah data

x_i = nilai aktual

\hat{x}_i = nilai prediksi

b. *Mean Squared Error* (MSE)

MSE mengukur rata-rata dari kuadrat selisih antara nilai aktual dan prediksi (Erdianita dkk., 2023). karena menggunakan kuadrat error, metrik ini sangat sensitif terhadap kesalahan besar.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (2.4)$$

Keterangan:

n = jumlah data

x_i = nilai aktual

\hat{x}_i = nilai prediksi

c. *Root Mean Square Error* (RMSE)

RMSE digunakan untuk mengukur rata-rata kuadrat dari kesalahan prediksi, kemudian diakarkan. Karena menggunakan kuadrat dari error, RMSE lebih sensitif terhadap nilai prediksi yang jauh berbeda dari nilai aktual RMSE merupakan akar kuadrat dari rata-rata kuadrat nilai aktual dan nilai prediksi (Ihzaniah dkk., 2023).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2} \quad (2.5)$$

Keterangan:

n = jumlah data

x_i = nilai aktual

\hat{x}_i = nilai prediksi

d. *Mean Absolute Percentage Error (MAPE)*

MAPE mengukur kesalahan prediksi dalam bentuk persentase terhadap nilai aktual (Ihzaniah dkk., 2023). Metrik ini membantu memahami seberapa besar kesalahan yang dibuat model dibandingkan dengan nilai sebenarnya. MAPE sangat berguna ketika ingin melihat performa model dalam skala yang relatif.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100\% \quad (2.6)$$

Keterangan:

n = jumlah data

x_i = nilai aktual

\hat{x}_i = nilai prediksi

e. Koefisien Determinasi (R^2 Score)

R^2 Score, atau koefisien determinasi, mengukur proporsi variansi dari nilai target yang dapat dijelaskan oleh fitur-fitur independen dalam model (Ihzaniah dkk., 2023). Semakin tinggi nilai R^2 , maka semakin baik model dalam menjelaskan variasi data.

$$R^2 = 1 - \frac{SSR}{SST} \quad (2.7)$$

Dengan:

- SSR = jumlah kuadrat sisa (residual)
- SST = jumlah kuadrat total

R^2 bernilai antara 0 dan 1. Nilai mendekati 1 menunjukkan bahwa model mampu menjelaskan sebagian besar variasi data, sedangkan nilai mendekati 0 menunjukkan bahwa model tidak dapat menjelaskan variansi data secara efektif.

2.2. Penelitian Terkait

2.2.1 State of The Art

Terdapat beberapa penelitian yang dilakukan menggunakan penelitian terkait dengan kasus serupa, berikut adalah penelitian tersebut terdapat pada Tabel 2.1

Tabel 2. 1 Penelitian Terkait

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
1.	(Lubis Halim dkk., 2024)	Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode <i>K-Nearest Neighbor</i>	<i>K-Nearest Neighbor</i>	Algoritma KNN digunakan untuk mengidentifikasi pola penjualan produk elektronik berdasarkan harga dan jumlah terjual, dengan akurasi prediksi mencapai 92%, menunjukkan keandalan tinggi dalam meramalkan produk terlaris.
2.	(Hia Abadi Sejahtera & Rahmawati, 2022)	Penerapan <i>K-Nearest Neighbor Regression</i> untuk Memprediksi Penjualan Tas Branded di Toko Shan Bogor	<i>K-Nearest Neighbor</i>	Penelitian ini, KNN diterapkan untuk mendeteksi pola dari data penjualan elektronik dengan mempertimbangkan variabel harga dan jumlah terjual, dan menghasilkan akurasi

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				<p>prediksi sebesar 92%, menunjukkan keandalannya dalam mengidentifikasi produk terlaris di masa depan.</p>
3.	(Aisha Alfani dkk., 2021)	Prediksi Penjualan Produk Unilever Menggunakan Metode <i>K-Nearest Neighbor</i>	<i>K-Nearest Neighbor</i>	<p>Hasil penelitian menunjukkan akurasi klasifikasi antara 40% hingga 86,66%, mencerminkan efektivitas sekaligus variabilitas metode K-NN dalam memprediksi penjualan. Penelitian ini menggunakan data tahun 2017–2019, sehingga mungkin belum merepresentasikan tren terkini. Selain itu, penggunaan data pelatihan dan pengujian yang terbatas (60 dan 20 data) dapat membatasi generalisasi hasil</p>

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				pada skala atau konteks yang lebih luas.
4.	(Rozi dkk., 2023)	Penerapan Machine Learning Untuk Prediksi Harga Saham PT.Teknologi Telekomunikasi Indonesia Tbk Menggunakan Algoritma <i>K-Nearest Neighbors</i>	<i>K-Nearest Neighbor</i>	Penelitian ini mengembangkan model prediksi harga saham menggunakan algoritma KNN dengan akurasi 85,8%. Model ini dirancang untuk membantu investor mengambil keputusan di tengah ketidakstabilan pasar, dengan memprediksi arah pergerakan harga saham secara andal.
5.	(Pangwijaya dkk., 2019)	Implementasi Algoritma Neural Network untuk Mendeteksi Daging Sapi dan Daging Kuda	<i>Neural Network</i>	Penelitian ini mengembangkan sistem klasifikasi berbasis analisis citra untuk membedakan daging sapi dan daging kuda menggunakan algoritma <i>neural network</i> . dengan mengambil nilai piksel dari

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				gambar sebagai data, sistem mencapai akurasi 76,67% untuk daging sapi dan 73,33% untuk daging kuda.
6.	(PurnaWibawa dkk., 2022)	Komparasi Algoritme <i>K-Nearest Neighbors</i> Dan <i>Support Vector Machines</i> Dalam Prediksi Layanan Produk ICONNET	<i>K-Nearest Neighbors (KNN)</i> Dan <i>Support Vector Machines (SVM)</i>	Penelitian ini membandingkan kinerja algoritma KNN dan SVM dalam memprediksi layanan ICONNET milik ICON+. Hasil menunjukkan bahwa KNN dengan Chebyshev Distance memberikan akurasi lebih tinggi (99,2%) dibanding SVM (94,4%) pada data training sebanyak 2.565 dan data testing 641, terutama dalam aspek akurasi dan AUC.
7.	(Bastian Anggun dkk., 2024)	Implementasi Algoritma <i>Support Vector Regression</i> dan <i>Linear</i>	<i>Support Vector Regression</i> dan <i>Linear</i>	Penelitian ini membandingkan efektivitas <i>Support Vector Regression</i> (SVR) dan <i>Multiple Linear</i>

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
		<i>regression</i> Untuk Prediksi Harga Rumah	<i>Linear regression</i>	<p><i>Regression</i> (MLR) dalam memprediksi harga rumah menggunakan 547 data dari Kecamatan Telukjambe Timur yang diperoleh melalui web scraping. SVR dioptimalkan dengan GridSearchCV dan menghasilkan kinerja lebih baik dibanding MLR, dengan RMSE sebesar 234.257 dan MAPE 21%. Hasil ini menegaskan bahwa SVR lebih akurat dalam prediksi harga rumah.</p>
8.	(Ardiansyah Firman, 2024)	Algoritma <i>Random Forest</i> Untuk Prediksi Harga Properti	<i>Random Forest</i>	Penelitian ini mengeksplorasi penggunaan <i>Random Forest</i> untuk memprediksi harga properti dengan hasil akurat, ditunjukkan oleh MAE dan R ² sekitar 0,89. Keakuratan model didukung oleh pra-

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				pemrosesan data dan optimalisasi parameter yang efektif. Meski membutuhkan komputasi tinggi dan tantangan integrasi sistem, Random Forest terbukti mampu mendukung pengambilan keputusan investasi dan meningkatkan efisiensi pasar properti.
9.	(Rangkuti Fauzi, 2024)	Anilisis dan Perbandingan Algoritma <i>Naïve Bayes</i> dan <i>C4.5</i> Dalam Klasifikasi Daging Sapi Terbaik	<i>Naïve Bayes</i> dan <i>C4.5</i>	Penelitian ini membandingkan algoritma <i>Naïve Bayes</i> dan <i>C4.5</i> dalam mengklasifikasikan kualitas daging sapi segar dan yang mengandung bahan tambahan. <i>C4.5</i> bekerja dengan mencari nilai gain tertinggi, sedangkan <i>Naïve Bayes</i> menghitung probabilitas berbobot pada atribut. Berdasarkan hasil

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				<p><i>Confusion Matrix, Naïve Bayes</i> menunjukkan akurasi klasifikasi yang lebih tinggi dibandingkan C4.5.</p>
10.	(Widaningsih Sri, 2019)	Perbandingan Metode <u>Data Mining</u> Untuk Prediksi Nilai dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, <i>Naïve Bayes</i> , KNN, dan SVM	C4.5, <i>Naïve Bayes</i> , KNN, dan <i>SVM</i>	<p>Penelitian ini membandingkan metode <i>data mining</i>—C4.5, <i>Naïve Bayes</i>, KNN, dan <i>SVM</i>—untuk memprediksi tingkat kelulusan dan IPK mahasiswa Teknik Informatika Universitas Suryakancana. Menggunakan pendekatan KDD pada data 2008–2013, fokus analisis mencakup jenis kelamin dan IPK semester 3–6. Hasil menunjukkan <i>Naïve Bayes</i> memiliki akurasi tertinggi sebesar 76,79% dalam memprediksi kelulusan tepat</p>

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				waktu dan IPK $\geq 3,00$, yang mendukung pemanfaatan data akademik untuk pengambilan keputusan yang lebih efektif.
11	(Rahmani dkk., 2024)	<i>Using Probabilistic Machine Learning Methods to Improve Beef Cattle Price Modeling and Promote Beef Production Efficiency and Sustainability in Canada</i>	<i>SVR, Addabost</i>	Penelitian ini menunjukkan bahwa algoritma machine learning multivariat, khususnya <i>Random Forest</i> dan <i>Adaboost</i> , memberikan performa prediksi terbaik dalam memodelkan harga sapi potong di Alberta, Kanada, dibandingkan dengan model univariat seperti ARIMA dan SARIMA. Dengan data bulanan dari 2005 hingga 2023, pendekatan probabilistik diterapkan untuk menangani ketidakpastian prediksi.

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				<p>Hasilnya membuktikan bahwa model multivariat yang dioptimasi dan dipadukan dengan pendekatan probabilistik mampu meningkatkan akurasi prediksi dan mendukung efisiensi serta keberlanjutan industri sapi potong di Kanada.</p>
12	(Hafid, 2023)	Penerapan <i>K-Fold Cross Validation</i> untuk Menganalisis Kinerja Algoritma <i>K-Nearest Neighbor</i> pada Data Kasus Covid-19 di Indonesia	<i>K-Nearest Neighbor</i>	<p>Penelitian ini mengevaluasi kinerja algoritma <i>K-Nearest Neighbor</i> (K-NN) dalam memprediksi kasus Covid-19 di Indonesia menggunakan metode <i>30-Fold Cross Validation</i>. Dengan nilai $k=5$, model K-NN mencapai akurasi sebesar 68,65% dan nilai kappa 0,5123, menunjukkan kemampuan prediksi yang cukup baik</p>

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				<p>serta tingkat kesepakatan yang moderat. Hasil ini memberikan wawasan mengenai efektivitas K-NN dalam analisis data Covid-19 dan dapat menjadi dasar pengembangan model prediksi yang lebih akurat di masa mendatang.</p>
13	(Muhamad Malik Matin, 2023)	<i>Hyperparameter Tuning Menggunakan GridsearchCV pada Random Forest untuk Deteksi Malware</i>	<i>Random Forest</i>	<p>Penelitian ini mengevaluasi peningkatan performa algoritma <i>Random Forest</i> dalam deteksi <i>malware</i> melalui penerapan <i>hyperparameter tuning</i> menggunakan <i>GridSearchCV</i>. Eksperimen menunjukkan bahwa parameter terbaik—di antaranya <i>criterion=entropy</i>,</p>

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				<p>max_depth=128,</p> <p>max_features=log2,</p> <p>min_samples_split=2, dan</p> <p>n_estimators=400—</p> <p>berhasil meningkatkan performa model, khususnya recall sebesar 0,37% serta akurasi dan F1-score masing-masing sebesar 0,19%. Hasil ini menunjukkan bahwa tuning <i>hyperparameter</i> mampu mengoptimalkan kinerja Random Forest dalam klasifikasi <i>malware</i>.</p>
14	(Akbar & Kusumodestoni, 2020)	<i>Optimization of k value and lag parameter of k-nearest neighbor algorithm on the prediction of hotel occupancy rates</i>	<i>k-nearest neighbor</i>	<p>Penelitian ini mengkaji prediksi tingkat hunian hotel menggunakan algoritma <i>k-Nearest Neighbor</i> (kNN) dengan optimalisasi parameter lag dan nilai k. Data histori hunian disusun sebagai data <i>supervised</i>, dan</p>

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				<p>model divalidasi menggunakan <i>10-fold cross-validation</i> dengan variasi <i>k</i> dari 1 hingga 30. Hasil eksperimen menunjukkan bahwa interval lag optimal berada pada 14–17, sedangkan nilai <i>k</i> optimal berada pada 5–13 untuk prediksi tingkat hunian hotel 1, 3, 6, 9, dan 12 bulan ke depan. Temuan ini memberikan dasar pemodelan prediktif yang lebih akurat dalam pengelolaan bisnis perhotelan.</p>
15	(Muhammad dkk., 2024)	Perbandingan Algoritma <i>Decision Tree</i> dan <i>K-Nearest Neighbor</i> untuk Klasifikasi	<i>Decision Tree</i> dan <i>K-Nearest Neighbor</i>	<p>Penelitian ini membandingkan algoritma <i>Decision Tree</i> dan <i>K-Nearest Neighbor</i> (KNN) dalam mendekripsi serangan pada</p>

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
		Serangan Jaringan IoT		<p>jaringan IoT menggunakan dataset Edge-IIoTset. Tujuan utamanya adalah memperoleh hasil klasifikasi terbaik dalam skenario klasifikasi biner dan multikelas. Hasil evaluasi menunjukkan bahwa algoritma <i>Decision Tree</i> memiliki performa lebih unggul dibandingkan KNN, baik dalam klasifikasi biner maupun multikelas. Pada klasifikasi biner, <i>Decision Tree</i> mengungguli KNN dengan selisih presisi, <i>recall</i>, <i>F1-score</i>, dan akurasi masing-masing sebesar 0,15; 0,18; 0,17; dan 0,08. Sementara itu, pada klasifikasi multikelas, selisih performa berturut-turut</p>

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				mencapai 0,26; 0,20; 0,22; dan 0,23.
16	(Azis dkk., 2020)	Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung	<i>K-Nearest Neighbor</i>	Penelitian ini mengevaluasi performa algoritma <i>K-Nearest Neighbor</i> (K-NN) dengan metode cross-validation dalam klasifikasi penyakit kardiovaskular menggunakan dataset dari UCI <i>Machine Learning Repository</i> . Dataset terdiri dari 1000 data pasien dengan 11 atribut, mencakup penderita dan non-penderita penyakit kardiovaskular. Simulasi dilakukan dengan rasio data 20:80, 50:50, dan 80:20, serta variasi nilai K (2–900) menggunakan 10-fold cross-validation. Hasil terbaik diperoleh pada rasio 20:80 dengan akurasi 87%,

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				presisi 87%, recall 97%, dan f-measure 92% pada nilai K=3. Temuan ini menunjukkan bahwa K-NN mampu memberikan performa prediksi yang baik dalam mendeteksi penyakit kardiovaskular.
17	(Mutiara, 2015)	Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4	K-Nearest Neighbor	Penelitian ini memanfaatkan data akademik Program Studi Ilmu Komputer FMIPA Unlam untuk memprediksi kelulusan tepat waktu mahasiswa menggunakan metode <i>k-Nearest Neighbor</i> (kNN). Pemilihan nilai k dilakukan dengan metode <i>k-Fold Cross Validation</i> dan uji akurasi untuk menentukan nilai k optimal, yang didapatkan sebesar 5 dengan akurasi 80,00%. Nilai k=5

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				kemudian digunakan pada algoritma kNN untuk memprediksi kelulusan tepat waktu berdasarkan IP hingga semester 4.
18	(Nurfauzan & Fatimah, 2022)	Implementasi Algoritma <i>K-Nearest Nearest Neighbors Regression</i> Dalam Memprediksi Harga Saham	<i>K-Nearest Neighbor</i>	Penelitian ini membangun model prediksi harga penutupan saham PT Bank Negara Indonesia Tbk menggunakan algoritma k-nearest neighbors regression dengan empat variasi fitur indikator, yaitu satu hingga empat fitur yang meliputi <i>open</i> , <i>high</i> , <i>low</i> , dan <i>close</i> . Dataset yang digunakan berasal dari Yahoo Finance periode Mei 2013 hingga Juli 2022 sebanyak 2274 data. Hasil menunjukkan bahwa model dengan satu indikator (fitur <i>close</i>) memberikan

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				<p>performa terbaik dengan nilai <i>root mean squared error</i> sebesar 169 dan akurasi mencapai 98,8% menggunakan pembagian data 70% untuk pelatihan dan 30% untuk pengujian.</p>
19	(Nanja Purwanto, 2015)	Metode <i>K-Nearest Neighbor</i> Berbasis <i>Forward Selection</i> Untuk Prediksi Harga Komoditi Lada	<i>K-Nearest Neighbor</i>	<p>Penelitian ini mengembangkan model prediksi komoditi lada dengan menggabungkan algoritma <i>K-Nearest Neighbor</i> (KNN) dan metode seleksi atribut <i>forward selection</i>. Evaluasi menggunakan data time series lada hitam dan lada putih menunjukkan bahwa KNN berbasis <i>forward selection</i> memberikan kinerja terbaik dibandingkan KNN dengan <i>backward elimination</i></p>

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
				dan SVM berbasis seleksi atribut.
20	(Shidiq, 2021)	Penerapan Metode <i>K-Nearest Neighbor</i> (KNN) Untuk Menentukan Ikan Cupang Dengan Ekstraksi Fitur Ciri Bentuk Dan Canny	<i>K-Nearest Neighbor</i> (KNN)	Penelitian ini mengaplikasikan metode <i>K-Nearest Neighbor</i> (KNN) dan deteksi tepi menggunakan metode Canny untuk menentukan jenis ikan betta. Metode Canny yang dikenal optimal dalam deteksi tepi dengan tingkat kesalahan rendah dikombinasikan dengan KNN menghasilkan akurasi yang cukup tinggi, yaitu rata-rata 68,57% untuk data pelatihan sebanyak 70 ekor dan 70% untuk data uji sebanyak 20 ekor ikan betta.
21	(Kusuma & Sasongko, 2023)	Optimasi <i>K-Nearest Neighbor</i> dengan <i>Grid Search CV</i>	<i>K-Nearest Neighbor</i> (KNN)	Penelitian ini mengoptimalkan nilai K pada algoritma KNN menggunakan <i>Grid Search</i>

No	Nama, Tahun Peneliti	Judul	Metode	Hasil Penelitian
		pada Prediksi Kanker Paru-Paru		CV untuk klasifikasi kanker paru-paru. Hasilnya menunjukkan bahwa nilai $K=3$ memberikan performa terbaik dengan akurasi 96%, <i>precision</i> 97%, <i>recall</i> 96%, dan <i>F1-score</i> 96%, menegaskan keunggulan kombinasi KNN dan tuning parameter otomatis dalam prediksi medis.

2.2.2 Matriks Penelitian

Tabel 2. 2 Matriks Penelitian

No	Penulis	Judul	Model/Algoritma						
			RF	C4.5	Naïve Bayes	ELM	SVM	KNN	Neural Network
1	(Lubis Halim dkk., 2024)	Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode <i>K-Nearest Neighbor</i>						√	
2	(Hia Abadi Sejahtera & Rahmawati, 2022)	Penerapan <i>K-Nearest Neighbor Regression</i> untuk Memprediksi Penjualan Tas Branded di Toko Shan Bogor						√	
3	(Aisha Alfani dkk., 2021)	Prediksi Penjualan Produk Unilever Menggunakan Metode <i>K-Nearest Neighbor</i>						√	
4	(Rozi dkk., 2023)	Penerapan Machine Learning Untuk Prediksi Harga Saham PT.Teknologi Indonesia Tbk Menggunakan						√	

No	Penulis	Judul	Model/Algoritma					
			RF	C4.5	Naïve Bayes	ELM	SVM	KNN
		Algoritma <i>K-Nearest Neighbors</i>						
5	(Pangwijaya dkk., 2019)	Implementasi Algoritma <i>Neural Network</i> untuk Mendeteksi Daging Sapi dan Daging Kuda						✓
6	(PurnaWibawa dkk., 2022)	Komparasi Algoritme <i>K-Nearest Neighbors</i> Dan <i>Support Vector Machines</i> Dalam Prediksi Layanan Produk ICONNET				✓	✓	
7	(Bastian Anggun dkk., 2024)	Implementasi Algoritma <i>Support Vector Regression</i> dan <i>Linear regression</i> Untuk Prediksi Harga Rumah						✓
8	(Ardiansyah Firman, 2024)	Algoritma <i>Random Forest</i> Untuk Prediksi Harga Properti	✓					
9	(Rangkuti Fauzi, 2024)	Anilisis dan Perbandingan Algoritma <i>Naïve Bayes</i> dan <i>C4.5</i> Dalam Klasifikasi		✓	✓			

No	Penulis	Judul	Model/Algoritma					
			RF	C4.5	Naïve Bayes	ELM	SVM	KNN
		Daging Sapi Terbaik						
10	(Widaningsih Sri, 2019)	Perbandingan Metode <u>Data Mining</u> Untuk Prediksi Nilai dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, <i>Naïve Bayes</i> , KNN, dan SVM		√	√		√	√
11	(Rahmani dkk., 2024)	<i>Using Probabilistic Machine Learning Methods to Improve Beef Cattle Price Modeling and Promote Beef Production Efficiency and Sustainability in Canada</i>					√	
12	(Hafid, 2023)	Penerapan <i>K-Fold Cross Validation</i> untuk Menganalisis Kinerja Algoritma <i>K-Nearest Neighbor</i> pada Data Kasus Covid-						√

No	Penulis	Judul	Model/Algoritma					
			RF	C4.5	Naïve Bayes	ELM	SVM	KNN
		19 di Indonesia Hafid H						
13	(Muhamad Malik Matin, 2023)	<i>Hyperparameter Tuning Menggunakan GridsearchCV pada Random Forest untuk Deteksi Malware</i>						
14	(Akbar & Kusumodestoni, 2020)	<i>Optimization of k value and lag parameter of k-nearest neighbor algorithm on the prediction of hotel occupancy rates</i>						√
15	(Muhammad dkk., 2024)	Perbandingan Algoritma <i>Decision Tree</i> dan <i>K-Nearest Neighbor</i> untuk Klasifikasi Serangan Jaringan IoT	√					
16	(Azis dkk., 2020)	Performa Klasifikasi K-NN dan <i>Cross Validation</i> Pada Data Pasien Pengidap Penyakit Jantung						√
17	(Mutriara, 2015)	Penerapan K-Optimal Pada					√	

No	Penulis	Judul	Model/Algoritma					
			RF	C4.5	Naïve Bayes	ELM	SVM	KNN
		Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4						
18	(Nurfauzan & Fatimah, 2022)	Implementasi Algoritma <i>K-Nearest Neighbors Regression</i> Dalam Memprediksi Harga Saham				✓		
19	(Nanja & Purwanto, 2015)	Metode <i>K-Nearest Neighbor</i> Berbasis <i>Forward Selection</i> Untuk Prediksi Harga Komoditi Lada				✓		
20	(Shidiq, 2021)	Penerapan Metode <i>K-Nearest Neighbor</i> (KNN) Untuk Menentukan Ikan Cupang Dengan Ekstraksi Fitur Ciri Bentuk Dan Canny				✓		
21	(Kusuma & Sasongko, 2023)	Optimasi <i>K-Nearest Neighbor</i> dengan <i>Grid</i>					✓	

No	Penulis	Judul	Model/Algoritma					
			RF	C4.5	Naïve Bayes	ELM	SVM	KNN
		<i>Search CV pada Prediksi Kanker Paru-Paru</i>						

2.3. Kebaruan Penelitian

Berdasarkan Matriks Penelitian Terkait diatas, Kebaruan dari penelitian ini terletak pada penerapan algoritma *K-Nearest Neighbors* (KNN) untuk memprediksi harga jual sapi dengan pendekatan regresi dan pembobotan jarak (distance-weighted), yang belum ditemukan dalam penelitian-penelitian sebelumnya. Berdasarkan perbandingan dengan studi (Rahmani dkk., 2024) dalam jurnal *Sustainability*, model-model seperti *Random Forest*, *AdaBoost*, dan *SARIMAX* telah digunakan untuk prediksi harga ternak, namun belum ada yang mengeksplorasi potensi KNN. Pendekatan berbasis KNN masih jarang digunakan, padahal memiliki keunggulan dalam kesederhanaan, fleksibilitas, serta performa yang kompetitif jika dikombinasikan dengan teknik validasi silang dan *tuning* parameter seperti *GridSearchCV* (Kusuma & Sasongko, 2023). Selain itu, dari beberapa penelitian (Lubis et al., 2024; Hia & Rahmawati, 2022) menunjukkan bahwa KNN mampu memberikan akurasi tinggi dalam prediksi penjualan produk, membuktikan efektivitasnya untuk menangkap pola data historis tanpa asumsi distribusi data tertentu, yang sangat relevan untuk kasus prediksi harga sapi.