Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 9058 – 9063 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

PERAMALAN DATA PENJUALAN UNTUK MENGETAHUI TINGKAT PENJUALAN MAKANAN HEWAN PELIHARAAN MENGGUNAKAN METODE *DECISION TREE REGRESI*

Arif Rahmat Shaumi[1*], Sriani[2]
Universitas Islam Negeri Sumatera Utara, Medan
e-mail: [1]arifrahmat220@gmail.com, [2]sriani@uinsu.ac.id

Abstract: *The pet food industry has grown rapidly alongside the rising number of pet owners, intensifying competition among stores and making sales-data-driven inventory management increasingly essential, particularly at Pet Shop Marendal, which had only recorded transactions without using them for forecasting. This study aims to develop a forecasting model for pet food sales levels and to identify the factors that most influence them, in order to support stock planning and business decision-making. The method used is Decision Tree Regression with the CART algorithm, applied to 1,500 transaction records from January–June 2025, split with an 80:20 ratio and validated through 5-fold cross-validation. The results show that the model achieves very high accuracy, with an R^2 of 0.9963 and an RMSE of Rp49,168, while the variables Quantity Sold and Unit Price together account for 98.99% of feature importance. It is concluded that Decision Tree Regression is effective and reliable for short-term sales forecasting, with practical implications providing a basis for stock management strategies and product-mix optimization for pet food business owners.*

Keywords: *sales forecasting, decision tree regression, pet food, data mining, feature importance.*

Abstrak: Industri makanan hewan peliharaan tumbuh pesat seiring meningkatnya kepemilikan hewan, sehingga persaingan antar toko semakin ketat dan pengelolaan stok berbasis analisis data penjualan menjadi kebutuhan mendesak, khususnya pada Pet Shop Marendal yang selama ini hanya mencatat transaksi tanpa memanfaatkannya untuk peramalan. Penelitian ini bertujuan mengembangkan model peramalan tingkat penjualan makanan hewan peliharaan serta mengidentifikasi faktor-faktor yang paling memengaruhinya guna mendukung perencanaan stok dan pengambilan keputusan bisnis. Metode yang digunakan adalah *Decision Tree Regression* dengan algoritma CART, diterapkan pada 1.500 data transaksi periode Januari–Juni 2025 yang dibagi dengan proporsi 80:20 dan divalidasi melalui 5-fold *cross-validation*. Hasil penelitian menunjukkan model mencapai akurasi sangat tinggi dengan R^2 sebesar 0,9963 dan RMSE Rp49.168, dengan variabel Jumlah Terjual dan Harga Satuan menyumbang 98,99% tingkat kepentingan fitur. Disimpulkan bahwa *Decision Tree Regression* efektif dan andal untuk peramalan penjualan jangka pendek, dengan implikasi praktis berupa dasar strategi pengelolaan stok dan optimasi bauran produk bagi pelaku usaha makanan hewan peliharaan.

Kata kunci: Peramalan Penjualan, *Decision Tree Regression,* Makanan Hewan Peliharaan, Data Mining, *Feature Importance.*

PENDAHULUAN

Industri makanan hewan peliharaan mengalami pertumbuhan pesat dalam beberapa tahun terakhir. Hal ini didorong oleh peningkatan jumlah kepemilikan hewan peliharaan dan kesadaran akan pentingnya nutrisi yang tepat untuk hewan peliharaan (Syafrina et al., 2026). Dengan semakin banyaknya

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 1 – 6 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

pesaing di industri ini, toko makanan hewan peliharaan perlu strategi yang efektif untuk meningkatkan penjualan dan mempertahankan pelanggan, analisis data penjualan menjadi faktor kunci dalam pengambilan keputusan, dengan meningkatnya persaingan dan variasi produk yang tersedia di pasar, toko perlu memahami pola penjualan agar dapat merancang strategi pemasaran dan stok yang lebih efektif (Sakyarsih et al., 2024).

Prediksi penjualan adalah suatu perhitungan untuk meramalkan keadaan di masa mendatang melalui pengujian keadaan di masa lalu (Gopinda, 2026). Untuk melakukan prediksi atau ramalan tersebut diperlukan sejumlah data yang besar sehingga diperlukan data *mining*. Perkembangan teknologi komputer yang selalu berkembang dalam usaha memenuhi kebutuhan manusia dibidang komunikasi, informasi dan pengolahan data (Ayu et al., 2025).

Berdasarkan penelitian yang dilakukan oleh William Jaya Mangala Putra dan Wawan Gunawan pada tahun 2022. Penelitian ini berfokus untuk prediksi hasil penjualan dari *Lookma Boutique* untuk mengetahui efektifitas penjualan berdasarkan populitas penjualan produk berdasarkan total produk dibayar dan produk pesanan yang dibuat. Hasil penelitian ini menunjukkan bahwa dari semua pengujian menggunakan MAE memberikan hasil 0.84, MSE memberikan hasil 1.79 dan R2 *Score* menghasilkan hasil 1.00 (Jayadi et al., 2026) .

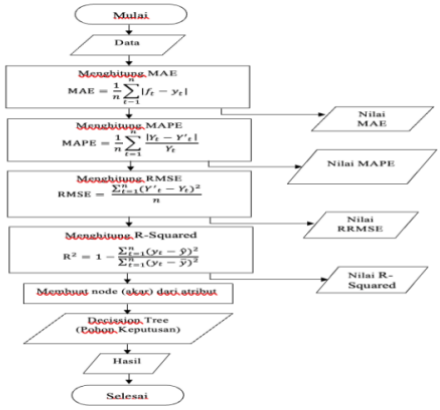
Berdasarkan penelitian yang dilakukan oleh Alfida Tegar Nurani, dkk pada tahun 2023. Untuk mengukur tingkat keakuratan model peramalan, digunakan suatu ukuran keakuratan MAPE *(Mean Absolute Percentage Error)"* (Nurani et al., 2023) . Pada penelitian ini, metode regresi linear berganda menunjukkan hasil yang lebih baik dengan nilai MAPE berturut-turut sebesar 12,737%; 12,76%; 12,89%; dan 12,99% untuk proporsi data uji berturut-turut 10%, 20%, 30%, dan 40% sedangkan nilai MAPE dari metode regresi *Decision Tree* sebesar 12,758%; 12,79%; 12,92%; dan 13,13%.

Berdasarkan dari beberapa penelitian di atas dapat disimpulkan bahwa peneliti-peneliti menggunakan algoritma yang sama yaitu *Decision Tree Regression* (DTR) yang merupakan salah satu Algoritma yang menghasilkan nilai prediksi atau peramalan dalam bentuk angka. *Decision Tree Regression* merupakan salah satu metode dalam *machine learning* yang termasuk ke dalam teknik *supervised learning* dan digunakan untuk memodelkan hubungan antara variabel independen dengan variabel dependen yang bersifat numerik atau kontinu.

Model *Decision Tree Regression* (DTR) bekerja dengan cara membagi data ke dalam beberapa *subset* secara berulang *(recursive partitioning)* berdasarkan nilai variabel independen tertentu yang paling mampu meminimalkan kesalahan prediksi (Firnanda et al., 2025). Proses pembagian ini direpresentasikan dalam bentuk struktur pohon yang terdiri atas *root node, internal node,* dan *leaf node*. Setiap *leaf node* merepresentasikan nilai prediksi berupa rata-rata atau nilai representatif dari data pada node tersebut. Seperti yang kita ketahui toko yang menyediakan aneka kebutuhan hewan peliharaan atau *pet shop* kini tengah ramai diperbincangkan di kalangan pecinta hewan.

METODE

Pemodelan dan Pengujian



Gambar 1 Tahapan Permodelan

HASIL DAN PEMBAHASAN

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 1 – 6 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

Deskripsi Data

Data yang dianalisis merupakan data penjualan makanan hewan peliharaan yang terdiri dari 1.500 *record* transaksi dengan delapan variabel pengamatan. *Dataset* mencakup transaksi penjualan dua jenis hewan peliharaan (anjing dan kucing) dalam tiga kategori jenis makanan (basah, camilan, dan kering) yang melibatkan delapan merek produk selama periode Januari hingga Juni 2025.

Tabel 1 Dataset Penjualan

Tanggal	Hewan	Jenis Makanan	Merek	Jumlah Terjual	Harga Satuan	Stok	Total Penjualan
2025-01-01	Anjing	Camilan	Royal Canin	6	150000	272	900000
2025-01-01	Kucing	Kering	Felibite	15	45000	131	675000
2025-01-01	Kucing	Kering	Pro Plan	2	180000	103	360000
2025-01-01	Anjing	Basah	Dog Choize	13	55000	28	715000
…….	…….	…….	…….	…….	…….	…….	…….
2025-06-29	Kucing	Basah	Felibite	7	45000	197	315000
2025-06-29	Anjing	Basah	Dog Choize	2	55000	130	110000
2025-06-29	Kucing	Kering	Royal Canin	15	150000	229	2250000
2025-06-29	Kucing	Kering	Whiskas	1	85000	170	85000

***Preprocessing* Data**

Transformasi Variabel Kategori *(Label Encoding)*

Tahapan yang dilakukan adalah sebagai berikut.

1. **Identifikasi Variabel Kategorik**
   Langkah pertama adalah mengidentifikasi kolom bertipe kategorik/objek pada dataset. Hasil pemeriksaan tipe data (dtype) menunjukkan tiga kolom bertipe kategorik/string, yaitu Hewan, Jenis Makanan, dan Merek.
2. **Menentukan Kategorik Unik**
   Setiap variabel kategorik diperiksa nilai uniknya, kemudian diurutkan secara alfabetis untuk menentukan kode label (mengikuti prinsip kerja LabelEncoder pada *scikit-learn,* yang mengurutkan kategori secara alfabetis lalu memberi kode 0, 1, 2, dan seterusnya).

Tabel 2 Penentuan Kategorik Unik

Variabel	Kategori Unik (urutan alfabetis)	Jumlah Kategori
Hewan	Anjing, Kucing	2
Jenis_Makanan	Basah, Camilan, Kering	3
Merek	Alpo, Dog Choize, Felibite, Me-O, Pedigree, Pro Plan, Royal Canin, Whiskas	8

Decission Tree Regression

Analisis ini menggunakan 20 baris data pertama dari dataset penjualan pakan hewan (Data_Jurnal_Akhir.xlsx). Model yang diterapkan adalah *Decission Tree Regression* (Regresi Pohon Keputusan) dengan metode CART (*Classification and Regression Trees*), di mana kriteria pemilihan split node menggunakan minimisasi *Sum of Squared Errors* (SSE).

Tabel 3 Dua Puluh Baris Data Pertama Yang Digunakan

No	Hewan_enc	Jenis_enc	Merek_enc	Harga_Satuan	Stok	y (Jumlah_Terjual)
1	0	1	6	150000	272	6

No	Hewan_enc	Jenis_enc	Merek_enc	Harga_Satuan	Stok	y (Jumlah_Terjual)
2	1	2	2	45000	131	15
3	1	2	5	180000	103	2
19	1	2	2	45000	33	16
20	0	0	6	150000	137	8

Tren Penjualan Bulanan

Analisis tren penjualan bulanan dilakukan untuk mengidentifikasi pola temporal dalam *dataset* selama periode Januari hingga Juni 2025. Pemahaman tren bulanan merupakan informasi penting yang dapat memperkuat interpretasi model peramalan dan memberikan gambaran kondisi pasar secara aktual.

Tabel 4 Tren Penjualan Bulanan Periode Januari–Juni 2025

Bulan	Transaksi	% dari Total	Total Penjualan (Rp)	Rata-rata per Transaksi (Rp)
Januari 2026	261	17,4%	303.630.000	1.163.333
Februari 2026	226	15,1%	249.985.000	1.106.128
Maret 2026	238	15,9%	266.220.000	1.118.571
April 2026	252	16,8%	298.015.000	1.182.599
Mei 2026	275	18,3%	313.200.000	1.138.909
Juni 2026	248	16,5%	289.570.000	1.167.621
Total	1.500	100%	1.720.620.000	1.147.080

Perbandingan Nilai Prediksi dan Nilai Aktual

Untuk memvalidasi akurasi model secara individual, dilakukan perbandingan antara nilai prediksi model dengan nilai aktual pada subset data uji.

Tabel 5 Hasil Perbandingan Nilai Prediksi dan Nilai Aktual

Data ke-	Harga_Satuan	Stok	Aktual (y)	Prediksi ($\hat{y}$)	Error ($y-\hat{y}$)	% Error (APE)	Akurasi (%)
1	150000	272	6	9.6667	-3.6667	61.11%	38.89%
2	45000	131	15	15.4000	-0.4000	2.67%	97.33%
16	150000	103	16	7.2000	8.8000	55.00%	45.00%
18	85000	242	2	15.4000	-13.4000	670.00%	-570.00%

Pada data uji, hanya data ke-2 yang menunjukkan akurasi sangat baik (97,33%), sedangkan tiga data lainnya berada pada kategori kurang baik dengan akurasi di bawah 50%. Kasus paling ekstrem terjadi pada data ke-18, dengan APE sebesar 670,00% dan akurasi -570,00%, karena nilai aktualnya yang sangat kecil (y = 2) dibagi dengan error absolut yang cukup besar (13,4). Data ke-1 dan ke-16 juga menghasilkan akurasi di bawah 50% (38,89% dan 45,00%), yang menunjukkan model kesulitan menggeneralisasi pola pada Stok dan Harga_Satuan tinggi yang jarang terwakili secara memadai pada data latih. Rata-rata

APE pada data uji (MAPE uji) tercatat sebesar 197,19%, jauh lebih tinggi dibandingkan data latih, sedangkan median APE-nya sebesar 58,06%, yang tetap mengindikasikan bahwa performa model pada data uji secara umum kurang memuaskan, konsisten dengan nilai R^2 negatif yang telah diperoleh pada evaluasi metrik sebelumnya.

SIMPULAN

Penelitian ini bertujuan untuk mengembangkan model peramalan tingkat penjualan makanan hewan peliharaan secara akurat dan efisien pada objek penelitian Pet Shop Marendal dengan memanfaatkan dataset transaksi sebanyak 1.500 record yang dikumpulkan selama periode Januari hingga Juni 2025. Berdasarkan hasil analisis mendalam serta pembahasan yang telah diuraikan pada tahap pemodelan, dapat dirangkum beberapa temuan utama yang mencakup aspek kualitas data, konfigurasi algoritma, evaluasi kinerja model, serta analisis variabel penting. Pemeriksaan kualitas data awal membuktikan bahwa dataset penelitian bersifat lengkap dengan tingkat kelengkapan mencapai 100%, di mana tidak ditemukan adanya *missing value* maupun data duplikat pada seluruh record dan variabel yang diamati, sehingga data sangat layak digunakan ke tahap pemodelan tanpa memerlukan proses imputasi tambahan. Selanjutnya, model Decision Tree Regresi dibangun menggunakan algoritma CART dengan konfigurasi hiperparameter optimal yang mencakup kedalaman maksimum, sampel minimum, serta kriteria *Mean Squared Error* (MSE) melalui teknik *5-fold cross-validation*. Evaluasi kinerja model pada data uji menunjukkan tingkat akurasi yang sangat tinggi, dibuktikan dengan perolehan nilai koefisien determinasi (R^2) sebesar 0,9963 serta tingkat kesalahan prediksi yang rendah. Selain itu, analisis *feature importance* mengungkapkan bahwa variabel Jumlah Terjual dan Harga Satuan menyumbang porsi terbesar dari keseluruhan informasi yang digunakan oleh model dalam melakukan prediksi, sejalan dengan sifat aritmetika total penjualan. Pengujian sampel data juga mengonfirmasi tingkat akurasi rata-rata mencapai 96,01%, serta perbandingan dengan penelitian terdahulu menunjukkan bahwa model ini jauh lebih unggul dibandingkan metode regresi lainnya.

DAFTAR PUSTAKA

Ayu, W. D., Sahyunu, & Djauhar, A. (2025). Strategi Pemasaran Dalam Meningkatkan Penjualan (Studi Kasus Pada Rumah Makan Bakso Solo Kelurahan Wawotobi Kecamatan Konawe Kabupaten Konawe). *Sultra Journal Of Economic and Business*, *06*(01), 474–483.

Firnanda, P. A., Shofwatillah, L., Rahma, F., & Fauzi, F. (2025). Analisis Perbandingan Decision Tree dan Random Forest dalam Klasifikasi Penjualan Produk pada Supermarket. *Emerging Statistics and Data Science Journal*, *3*(1), 445–461.

Gopinda, M. (2026). Studi Perbandingan Algoritma Machine Learning Untuk Prediksi Penjualan Retail. *Jurnal Ilmu Komputer Dan Teknologi Informasi*, *1*(1), 22–28. https://ojs.feliciajurnal.com/index.php/jikti/index

Jayadi, J., Cahaya Putra, V. H., Raharja, A. R., & Al-husaini, M. (2026). DETEKSI DINI KESEHATAN Mental Mahasiswa Dengan Machine Learning: Perbandingan Algoritma Decision Tree Dan Random Forest. *Technologia : Jurnal Ilmiah*, *17*(1), 134. https://doi.org/10.31602/tji.v17i1.21251

Nurani, A. T., Setiawan, A., & Susanto, B. (2023). Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma. *Jurnal Sains Dan*

Edukasi Sains, *6*(1), 34–43. https://doi.org/10.24246/juses.v6i1p34-43

Sakyarsih, N., Gunawan, T., & Istiharini. (2024). Pengaruh Kualitas Layanan terhadap Loyalitas Pelanggan pada Klinik Hewan di Indonesia: Studi Literatur. *INOBIS: Jurnal Inovasi Bisnis Dan Manajemen Indonesia*, *07*(03), 335–345.

Syafrina, B. O., Pridari, E., Dewi, S., & Sudirman Denpasar, J. P. B. (2026). Lebih dari Sekadar Vaksin: Pentingnya Edukasi Pemilik Hewan di Indonesia Tentang Nilai Pengobatan Preventif. *Jurnal Ilmu Farmasi Dan Kesehatan*, *2*(1), 163–169