

Optimasi Strategi Penjualan Produk Zaskia Melalui Integrasi Algoritma K-Nearest Neighbors (KNN) dan Regresi Linier

¹Ermatita, ²Belly Nagustria,

^{1,2}Universitas Sriwijaya, Palembang, Indonesia

¹ermatitaz@yahoo.com, ²bellyssimo130885@gmail.com

Article Info

Article history:

Received, 2025-01-25

Revised, 2025-01-28

Accepted, 2025-01-31

Kata Kunci:

optimasi_penjualan
k_nearest_neighbors
regresi_linier
prediksi_produk_terlaris
e_commerce

Keywords:

sales_optimization
k_nearest_neighbors
linear_regression
best_selling_product_prediction
e_commerce

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi performa model dalam mengklasifikasikan serta memprediksi jumlah penjualan produk berdasarkan sejumlah atribut. Algoritma K-Nearest Neighbors (KNN) digunakan untuk tugas klasifikasi dan menunjukkan performa yang baik dengan akurasi sebesar 91,22%, recall 92,96%, dan precision 90,18%. Hasil ini mengindikasikan bahwa model memiliki kemampuan generalisasi yang tinggi dalam mengenali pola penjualan. Untuk prediksi kuantitatif jumlah penjualan, digunakan model regresi linier dengan variabel bebas berupa harga normal, harga jual, rating, jumlah penilaian, dan jumlah favorit. Model regresi menghasilkan nilai koefisien determinasi (R^2) sebesar 0,78, yang menunjukkan bahwa 78% variabilitas jumlah penjualan dapat dijelaskan oleh variabel-variabel tersebut. Hasil analisis koefisien menunjukkan bahwa rating, jumlah penilaian, dan jumlah favorit memiliki pengaruh positif terhadap jumlah penjualan, sedangkan harga jual dan harga normal tidak berpengaruh signifikan. Nilai Root Mean Squared Error (RMSE) sebesar 33,09 menunjukkan tingkat kesalahan prediksi yang cukup rendah. Temuan ini menunjukkan bahwa model yang digunakan efektif dalam membantu analisis dan peramalan penjualan produk.

ABSTRACT

This study aims to evaluate the model's performance in classifying and predicting the number of product sales based on several attributes. The K-Nearest Neighbors (KNN) algorithm was used for the classification task and showed good performance with an accuracy of 91.22%, recall of 92.96%, and precision of 90.18%. These results indicate that the model has a high generalization ability in recognizing sales patterns. For quantitative prediction of the number of sales, a linear regression model is used with independent variables such as regular price, selling price, rating, number of ratings, and number of favorites. The regression model yielded a coefficient of determination (R^2) of 0.78, indicating that 78% of the variability of the sales amount can be explained by these variables. The coefficient analysis results show that the rating, number of ratings, and number of favorites have a positive influence on the sales amount, while the selling price and normal price have no significant effect. The Root Mean Squared Error (RMSE) value of 33.09 indicates a fairly low level of prediction error. These findings indicate that the model used is effective in helping analyze and forecast product sales.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.



Penulis Korespondensi:

Ermamtitita,

Universitas Sriwijaya

Email: ermatitaz@yahoo.com

1. PENDAHULUAN

Dalam era digital yang berkembang pesat, e-commerce telah menjadi salah satu platform utama dalam aktivitas perdagangan modern [1]–[3]. Perubahan perilaku konsumen yang semakin bergantung pada layanan daring mendorong pelaku usaha untuk mengembangkan strategi penjualan yang adaptif dan berbasis data. Salah satu

tantangan besar yang dihadapi adalah bagaimana mengidentifikasi produk-produk yang berpotensi menjadi terlaris di tengah persaingan yang ketat dan preferensi konsumen yang dinamis [2], [4]–[6].

Brand Zaskia Sungkar Jakarta merupakan salah satu pelaku bisnis fashion muslimah yang aktif memanfaatkan platform e-commerce seperti Shopee. Meskipun produk-produknya dikenal memiliki kualitas dan estetika yang tinggi, namun strategi penjualan berbasis data masih perlu dioptimalkan guna meningkatkan efisiensi pengelolaan stok, kampanye promosi, dan konversi penjualan. Di sinilah peran teknologi data mining menjadi sangat penting.

K-Nearest Neighbors (KNN) adalah salah satu metode klasifikasi sederhana namun efektif yang telah digunakan dalam berbagai studi untuk mengelompokkan produk berdasarkan kesamaan fitur [1], [7]–[10]. Sementara itu, regresi linier memiliki keunggulan dalam memodelkan hubungan antara variabel-variabel numerik, seperti rating, harga, dan jumlah penilaian, terhadap jumlah penjualan. Integrasi kedua metode ini diharapkan mampu memberikan pendekatan prediktif dan klasifikatif yang komprehensif dalam mendukung pengambilan keputusan strategis [9], [11]–[13].

Beberapa penelitian terdahulu telah menunjukkan potensi metode ini dalam konteks prediksi penjualan. Penelitian oleh Takale et al. (2022) menunjukkan bahwa regresi linier efektif digunakan untuk memprediksi nilai penjualan produk secara numerik, meskipun terbatas pada hubungan linier antar variabel [14], [15]. Sementara itu, Rahmadden et al. (2024) membuktikan bahwa algoritma KNN mampu mengklasifikasikan produk berdasarkan tingkat penjualan dengan akurasi yang cukup tinggi dalam lingkungan e-commerce [16]. Dalam studi lain, Gopalakrishnan et al. (2018) juga mengemukakan bahwa integrasi beberapa parameter seperti harga, ulasan, dan rating melalui regresi linier menghasilkan akurasi prediksi penjualan sebesar 83,8% [15].

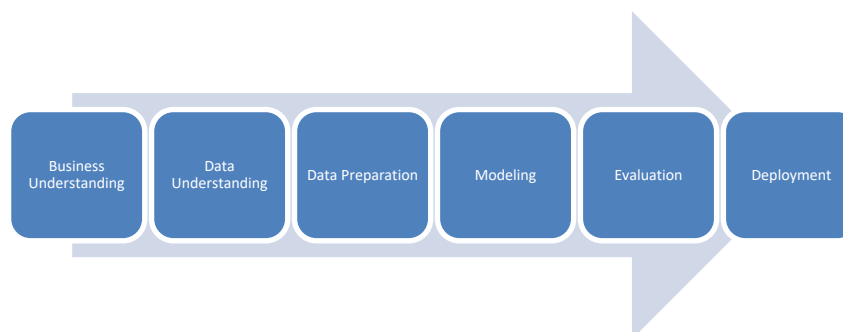
Penelitian oleh Wen et al. (2024) menggarisbawahi pentingnya pemilihan model prediktif yang tepat dalam konteks data penjualan ritel, di mana regresi dan model klasifikasi berbasis machine learning dapat digabungkan untuk meningkatkan akurasi [17]. Hal senada disampaikan oleh Dewi et al. (2022) yang menemukan bahwa teknik data mining, khususnya KNN, dapat digunakan untuk klasifikasi produk berdasarkan histori penjualan dan memberikan kontribusi dalam strategi pengelompokan produk secara otomatis [18].

Berangkat dari temuan-temuan tersebut, penelitian ini bertujuan untuk menerapkan dan mengintegrasikan metode KNN dan regresi linier dalam analisis data penjualan produk fashion muslimah milik Zaskia Sungkar Jakarta guna mengidentifikasi produk-produk terlaris secara efektif dan efisien.

2. METODE PENELITIAN

Penelitian ini merupakan jenis penelitian deskriptif kuantitatif yang menggunakan pendekatan data mining untuk mengeksplorasi dan memodelkan data penjualan produk Zaskia secara sistematis. Data yang digunakan berasal dari online shop Zaskia Sungkar Jakarta yang dijual melalui e-commerce Shopee. Pengumpulan data dilakukan secara langsung dari toko online dengan metode web scraping, kemudian disimpan dalam format spreadsheet Excel untuk pengolahan lebih lanjut.

Proses analisis menggunakan pendekatan CRISP-DM (Cross Industry Standard Process for Data Mining), yang terdiri dari tahapan berikut:



Gambar 1. Tahapan Penelitian

- a. Business Understanding. Tahapan ini digunakan untuk memahami tujuan bisnis sehingga dapat mengoptimalkan penjualan berdasarkan data historis penjualan.
- b. Data Understanding. Tahapan ini digunakan untuk mengevaluasi struktur data penjualan, seperti atribut harga, rating, penilaian, jumlah stok, dan jumlah terjual.

f. Deployment. Model digunakan untuk rekomendasi strategi penjualan bagi manajemen toko online.

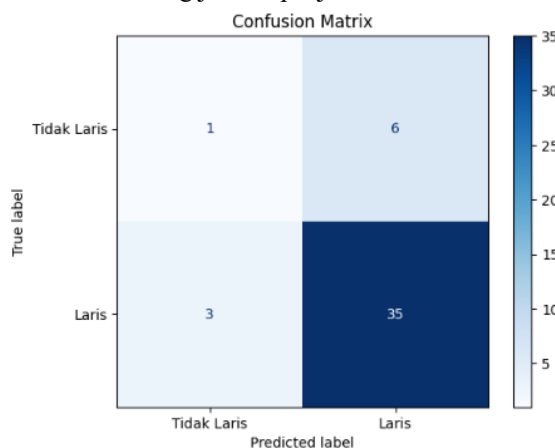
Penelitian ini menggunakan data penjualan produk fashion muslimah dari toko Zaskia Sungkar Jakarta yang diperoleh dari platform Shopee. Dataset mencakup produk dengan atribut-atribut utama yang meliputi: nama produk, harga normal, harga jual, jumlah penjualan, jumlah stok, rating, diskon, jumlah penilaian, dan jumlah favorit. Data tersebut digunakan untuk membangun model klasifikasi dan prediksi berbasis algoritma K-Nearest Neighbors (KNN) dan regresi linier.

contents	hinet-s0-w-full	src workspace-normal	truncate	text-sp10	text-sp10-iflex-none	iflex-none	truncate 2
https://schshttps://do/https://dowJade Bag Treasure ZS	599,000	Produkt Bar	Bebas Pen	https://de/c.4.9			37 Terjual
https://schshttps://do/https://dowSwan Blouse Another Chain ZS	449,000						
https://schshttps://do/https://dowClaire Bag Empire ZS	599,000	Bebas Pen		https://de/c.4.8			58 Terjual
https://schshttps://do/https://dowKavi Blouse Statement ZS	585,000	Produkt Bar	Bebas Pen	https://de/c.4.9			30 Terjual
https://schshttps://do/https://dowZS AMARYLLIS CALANTHA DRESS	599,000	Bebas Pen		https://de/c.4.9			76 Terjual
https://schshttps://do/https://dowSummer Blooms Books and Art ZS	469,000	Pilihi Lokal	Bebas Pen	https://de/c.4.9			303 Terjual
https://schshttps://do/https://dowMilla Blouse Black Another Chain ZS	449,000	Produkt Bar		https://de/c.5.0			6 Terjual
https://schshttps://do/https://dowMina Dress Another Chain ZS	659,000	Bebas Pen					
https://schshttps://do/https://dowDrew Blaise Purple Empire ZS	449,000	Bebas Pen		https://de/c.5.0			137 Terjual
https://schshttps://do/https://dowLyra Dress Purple Another ZS	939,000	Bebas Pen		https://de/c.4.9			21 Terjual
https://schshttps://do/https://dowMono Chain Scarf Another Chain ZS	285,000	Produkt Bar	Bebas Pen	https://de/c.4.9			22 Terjual
https://schshttps://do/https://dowTrixie Blouse Another Chain ZS	459,000	Produkt Bar					
https://schshttps://do/https://dowMora Oversized Shirt Treasure ZS	489,000	Produkt Bar		https://de/c.4.9			13 Terjual
https://schshttps://do/https://dowDemil Blouse Books and Art ZS	449,000	Bebas Pen		https://de/c.4.9			224 Terjual
https://schshttps://do/https://dowMilla Blouse Latte Another Chain ZS	445,000	Produkt Bar		https://de/c.5.0			4 Terjual
https://schshttps://do/https://dowMiss Dress Books and Art ZS	615,000	Pilihi Lokal	Bebas Pen	https://de/c.4.9			154 Terjual
https://schshttps://do/https://dowArisha Dress Sapphire ZS	629,000	Pilihi Lokal	Bebas Pen	https://de/c.4.9			222 Terjual
https://schshttps://do/https://dowBailey Blouse Another Chain ZS	449,000	Produkt Bar					
https://schshttps://do/https://dowBaila Knit Shirt Terracotta Books and Art ZS	449,000			https://de/c.4.9			80 Terjual
https://schshttps://do/https://dowSonaya Blouse Treasures ZS	499,000			https://de/c.5.0			36 Terjual
https://schshttps://do/https://dowBalgang Ring Empire ZS	149,000	Bebas Pen		https://de/c.4.8			50 Terjual

Gambar 2. Dataset

a. Hasil Klasifikasi Menggunakan Algoritma KNN

Model klasifikasi dibangun menggunakan algoritma K-Nearest Neighbors (KNN) dengan nilai $k = 3$, yang dipilih berdasarkan pengujian nilai k yang menghasilkan akurasi tertinggi. Tujuan dari klasifikasi ini adalah untuk mengelompokkan produk ke dalam dua kategori, yaitu "laris" dan "tidak laris", dengan target label ditentukan berdasarkan ambang jumlah penjualan.

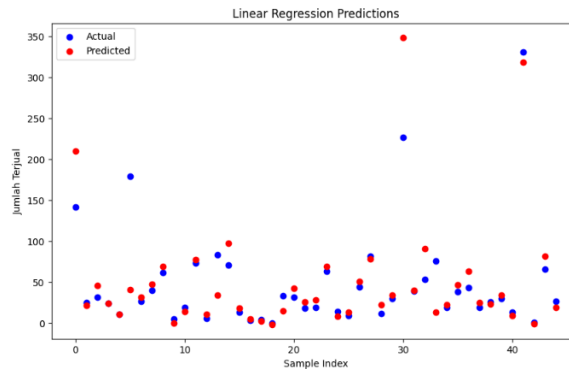


Gambar 3. Confusion matriks

Evaluasi model menunjukkan bahwa algoritma KNN menghasilkan akurasi klasifikasi sebesar 91,22%, dengan nilai recall sebesar 92,96% dan precision sebesar 90,18%. Nilai-nilai tersebut mengindikasikan

bahwa model memiliki kemampuan generalisasi yang baik dalam mengenali pola-pola penjualan produk berdasarkan atribut-atribut yang dimasukkan. Visualisasi performa model ditampilkan dalam bentuk confusion matrix (Gambar 1), yang menunjukkan dominasi klasifikasi benar (true positive dan true negative) dibandingkan dengan kesalahan klasifikasi.

b. Hasil Prediksi Menggunakan Regresi Linier



Gambar 4. Plot regresi

Untuk melakukan prediksi kuantitatif terhadap jumlah penjualan, model regresi linier digunakan. Variabel bebas yang digunakan dalam model ini meliputi harga normal, harga jual, rating, jumlah penilaian, dan jumlah favorit. Dari hasil regresi diperoleh model prediktif dengan nilai koefisien determinasi (R^2) sebesar 0,78, yang berarti bahwa sekitar 78% variabilitas jumlah penjualan dapat dijelaskan oleh variabel-variabel tersebut.

Koefisien regresi menunjukkan bahwa variabel rating ($\beta = 0,82$), jumlah penilaian ($\beta = 1,57$), dan jumlah favorit ($\beta = 0,05$) memiliki pengaruh positif terhadap jumlah penjualan, sementara harga jual dan harga normal tidak menunjukkan pengaruh signifikan. Evaluasi model menunjukkan nilai Root Mean Squared Error (RMSE) sebesar 33,09, yang menunjukkan bahwa prediksi penjualan rata-rata menyimpang sekitar 33 unit dari nilai aktual (Gambar 2).

4. KESIMPULAN

Model klasifikasi menggunakan algoritma K-Nearest Neighbors (KNN) menunjukkan performa yang baik dengan akurasi sebesar 91,22%, recall 92,96%, dan precision 90,18%. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat ketepatan dan sensitivitas yang tinggi terhadap pola penjualan produk. Visualisasi performa model ditampilkan melalui confusion matrix untuk memberikan gambaran detail terhadap hasil klasifikasi. Sementara itu, untuk prediksi kuantitatif jumlah penjualan, digunakan model regresi linier dengan variabel bebas berupa harga normal, harga jual, rating, jumlah penilaian, dan jumlah favorit. Model ini menghasilkan koefisien determinasi (R^2) sebesar 0,78, yang mengindikasikan bahwa 78% variabilitas jumlah penjualan dapat dijelaskan oleh variabel-variabel tersebut. Dari hasil analisis koefisien regresi, ditemukan bahwa rating ($\beta = 0,82$), jumlah penilaian ($\beta = 1,57$), dan jumlah favorit ($\beta = 0,05$) berpengaruh positif terhadap jumlah penjualan, sementara harga normal dan harga jual tidak menunjukkan pengaruh signifikan. Nilai Root Mean Squared Error (RMSE) sebesar 33,09 menunjukkan tingkat kesalahan prediksi yang masih berada dalam batas yang dapat diterima.

REFERENSI

- [1] W. Andriyani, M. Anshori, D. Normawati, R. S. Pradini, and ..., *Matematika Pada Kecerdasan Buatan*. books.google.com, 2024. [Online]. Available: https://books.google.com/books?hl=en%5C&lr=%5C&id=cEYhEQAAQBAJ%5C&oi=fnd%5C&pg=PP1%5C&dq=optimasi+penjualan+knn+regresi+linear%5C&ots=q4u7S1VN_a%5C&sig=VgZXgh2hdjY2TpByWX9PRhKnxQ4
- [2] R. F. Putra, I. R. Mukhlis, A. I. Datya, S. J. Pipin, F. Reba, and ..., *Algoritma Pembelajaran Mesin: Dasar, Teknik, dan Aplikasi*. books.google.com, 2024. [Online]. Available: https://books.google.com/books?hl=en%5C&lr=%5C&id=MX7-EAAAQBAJ%5C&oi=fnd%5C&pg=PA49%5C&dq=optimasi+penjualan+knn+regresi+linear%5C&ots=v7F7U_BAH-%5C&sig=fRd2qiKrQDpypc9KAWG6oAcfvKw

- [3] R. F. Putra, R. S. Y. Zebua, B. Budiman, P. W. Rahayu, and ..., *Data Mining: Algoritma dan Penerapannya*. books.google.com, 2023. [Online]. Available: <https://books.google.com/books?hl=en%5C&lr=%5C&id=zLHGEAAAQBAJ%5C&oi=fnd%5C&pg=PA22%5C&dq=optimasi+penjualan+knn+regresi+linear%5C&ots=tzPibspQF9%5C&sig=6tJa8RVEiWsAMJGaWeycmMidGdk>
- [4] S. T. Yahya, *Data mining*. books.google.com, 2022. [Online]. Available: https://books.google.com/books?hl=en%5C&lr=%5C&id=0J2mEAAAQBAJ%5C&oi=fnd%5C&pg=PR1%5C&dq=optimasi+penjualan+knn+regresi+linear%5C&ots=LQaFB5liql%5C&sig=ZdPwSF_EbxfKyyHRDOLbBAZj0M
- [5] T. Hidayat, "Data Mining untuk Meningkatkan Efisiensi dan Prediksi Produk Garmen Menggunakan Algoritma K-Nearest Neighbor di PT Mas Silueta," *Jurnal Kridatama Sains Dan Teknologi*, vol. 6, no. 01. Universitas Ma'arif Nahdlatul Ulama Kebumen, pp. 160–173, 2024. doi: 10.53863/kst.v6i01.1085.
- [6] N. R. Pambudi, *ANALISIS PREDIKSI PRODUK TERLARIS PADA SUPERMARKET BOERSA KAMPUS MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER DAN REGRESI ...*. eprints.amikompuwokoerto.ac.id, 2023. [Online]. Available: <https://eprints.amikompuwokoerto.ac.id/id/eprint/1837/>
- [7] G. Maulani, S. Kom, M. Kom, F. N. Hasan, S. Kom, and ..., *Machine Learning*. books.google.com, 2025. [Online]. Available: https://books.google.com/books?hl=en%5C&lr=%5C&id=RblPEQAAQBAJ%5C&oi=fnd%5C&pg=PR8%5C&dq=optimasi+penjualan+knn+regresi+linear%5C&ots=emlIuJs_PN%5C&sig=n4vA-X-89thgxnfnd8JJokn3JTW
- [8] A. I. Silitonga, *KECERDASAN BUATAN PADA BISNIS DIGITAL*. books.google.com, 2024. [Online]. Available: <https://books.google.com/books?hl=en%5C&lr=%5C&id=Yr4uEQAAQBAJ%5C&oi=fnd%5C&pg=PA1%5C&dq=optimasi+penjualan+knn+regresi+linear%5C&ots=bU1shGEp4I%5C&sig=QmpRkLjsflwqtuya6gEcipEp4UQ>
- [9] K. Kushariyadi, H. Apriyanto, Y. Herdiana, F. H. Asy'ari, and ..., *Artificial Intelligence: Dinamika Perkembangan AI Beserta Penerapannya*. books.google.com, 2024. [Online]. Available: <https://books.google.com/books?hl=en%5C&lr=%5C&id=lg0nEQAAQBAJ%5C&oi=fnd%5C&pg=PA44%5C&dq=optimasi+penjualan+knn+regresi+linear%5C&ots=Oj47T9fiW5%5C&sig=IIItnOqJVg6N16Q4NBUUp0McvW80>
- [10] I. D. Id, *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python*. books.google.com, 2021. [Online]. Available: https://books.google.com/books?hl=en%5C&lr=%5C&id=JvBPEAAAQBAJ%5C&oi=fnd%5C&pg=PA1%5C&dq=optimasi+penjualan+knn+regresi+linear%5C&ots=j-0lPGsgaw%5C&sig=SNrTihZAHF_DNdZKdoedB15j1Oc
- [11] S. Syam, Y. Tokoro, L. Judijanto, M. Garonga, F. M. Sinaga, and ..., *Data Mining: Teori dan Penerapannya dalam Berbagai Bidang*. books.google.com, 2024. [Online]. Available: <https://books.google.com/books?hl=en%5C&lr=%5C&id=hTAXEQAAQBAJ%5C&oi=fnd%5C&pg=PA62%5C&dq=optimasi+penjualan+knn+regresi+linear%5C&ots=hyIFj10xml%5C&sig=rpAjZQbSbfl8ChEZY8X3kyG0K1M>
- [12] F. Sulianta, *Basic Data Mining from A to Z*. books.google.com, 2023. [Online]. Available: <https://books.google.com/books?hl=en%5C&lr=%5C&id=JcLhEAAAQBAJ%5C&oi=fnd%5C&pg=PA1%5C&dq=optimasi+penjualan+knn+regresi+linear%5C&ots=VnCsPkYNro%5C&sig=542tKK1OzUMQrWacHpZ0WJxXJvA>
- [13] R. Swastika, S. Mukodimah, F. Susanto, M. Muslihudin, and ..., *IMPLEMENTASI DATA MINING (Clustering, Association, Prediction, Estimation, Classification)*. books.google.com, 2023. [Online]. Available: <https://books.google.com/books?hl=en%5C&lr=%5C&id=LsOqEAAAQBAJ%5C&oi=fnd%5C&pg=PA67%5C&dq=optimasi+penjualan+knn+regresi+linear%5C&ots=08zPtBL4Gw%5C&sig=95rymxOsCtPDUUVIi6clpaOg2B4>
- [14] S. Takale, T. Bhong, U. Dethe, and P. Gandhi, "Sales Prediction using Linear Regression," *J. Electron. Netw. Appl. Math.*, no. 25, pp. 62–71, Sep. 2022, doi: 10.55529/jecnam.25.62.71.
- [15] G. T., R. Choudhary, and S. Prasad, "Prediction of Sales Value in Online shopping using Linear Regression," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, Dec. 2018, pp. 1–6. doi: 10.1109/ccaa.2018.8777620.
- [16] N. Afiatuddin, M. T. Wicaksono, V. R. Akbar, R. Rahmadden, and D. Wulandari, "Komparasi Algoritma Machine Learning dalam Klasifikasi Kanker Payudara," *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 2, p. 889, Apr. 2024, doi: 10.30865/mib.v8i2.7457.

- [17] K. Y. Wen, M. H. Joseph, and V. Sivakumar, "Big Mart Sales Prediction using Machine Learning," *EAI Endorsed Trans. Internet Things*, vol. 10, Jun. 2024, doi: 10.4108/eetiot.6453.
- [18] R. Iriane, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Pangan Hewan Menggunakan Metode K-Nearest Neighbor," *KLIK Kaji. Ilm. Inform. dan Komput.*, 2023, [Online]. Available: <http://djournals.com/klik/article/view/683>