

---

## **PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK KLASIFIKASI RISIKO PRODUK KADALUARSA**

Syahprian

Program Studi Manajemen Informatika, Politeknik Ganesha Medan  
[syahprian29@gmail.com](mailto:syahprian29@gmail.com)

---

### **Abstrak**

Produk yang telah kadaluarsa adalah salah satu tantangan utama dalam pengelolaan inventaris di sektor ritel, terutama bagi usaha menengah yang belum menerapkan sistem pengelolaan stok berbasis data. Studi ini menawarkan penerapan algoritma K-Nearest Neighbor (KNN) untuk menentukan tingkat risiko kadaluarsa produk dalam tiga kategori: Aman, Waspada, dan Kritis. Data yang dianalisis merupakan data sekunder dari unit usaha menengah dengan atribut termasuk sisa waktu hingga kadaluarsa, persentase stok yang ada, dan tingkat perputaran barang. Pengujian dilakukan berbagai nilai  $k$  ( $k=1, 3, 5, 7, 9$ ) melalui metode validasi silang 10-fold. Hasil dari evaluasi menunjukkan bahwa konfigurasi terbaik dicapai pada  $k=5$  dengan akurasi sebesar 88.33%, precision 0.89, recall 0.88, dan F1-Score 0.89. Hasil ini menegaskan bahwa algoritma KNN dapat diterapkan secara efektif sebagai elemen dalam sistem pendukung keputusan untuk pengelolaan inventaris ritel.

*Kata Kunci—K-Nearest Neighbor, klasifikasi produk, kadaluarsa, inventori ritel, machine learning.*

### *Abstract*

Expired products are one of the main challenges in inventory management in the retail sector, especially for medium-sized businesses that have not implemented a data-driven stock management system. This study proposes the application of the K-Nearest Neighbor (KNN) algorithm to determine the risk level of product expiration in three categories: Safe, Alert, and Critical. The analyzed data are secondary data from medium-sized business units with attributes including remaining time to expiration, percentage of stock on hand, and turnover rate. Testing was carried out at various  $k$  values ( $k=1, 3, 5, 7, 9$ ) through a 10-fold cross-validation method. The results of the evaluation show that the best configuration is achieved at  $k=5$  with an accuracy of 88.33%, precision of 0.89, recall of 0.88, and F1-Score of 0.89. These results confirm that the KNN algorithm can be effectively implemented as an element in a decision support system for retail inventory management.

*Keywords—K-Nearest Neighbor, product classification, expiration, retail inventory, machine learning.*

---

## I. PENDAHULUAN

Pengelolaan inventaris yang efisien merupakan elemen penting bagi kelangsungan operasional bisnis ritel. Salah satu masalah besar yang sering muncul adalah kerugian yang disebabkan oleh barang yang melewati batas waktu jual sebelum berhasil terjual. Situasi ini tidak hanya menyebabkan kerugian, tetapi juga bisa mengurangi kepercayaan konsumen terhadap bisnis tersebut. Dalam konteks industri ritel di seluruh dunia, kerugian yang ditimbulkan oleh barang yang telah expired di Indonesia menimbulkan kerugian besar bagi pelaku usaha ritel [1].

Pendekatan machine learning, terutama melalui algoritma klasifikasi, telah terbukti sangat berguna untuk mendukung keputusan berbasis data di berbagai sektor bisnis. Novitadewi et al. [2] mengungkapkan bahwa metode KNN dapat dengan efektif mengelompokkan data penjualan produk dengan menggunakan atribut transaksi yang diambil dari sistem informasi yang sudah ada. Selain itu, Kusumah et al. [3] membuktikan bahwa KNN memberikan hasil yang memuaskan dalam memprediksi dan klasifikasi produk terlaris menggunakan pendekatan KDD pada data penjualan ritel.

K-Nearest Neighbor (KNN) dipilih karena cara penerapannya yang mudah, memiliki tingkat interpretasi yang tinggi, dan tidak memerlukan prasyarat mengenai distribusi data tertentu. Studi ini bertujuan untuk menerapkan KNN dalam menentukan klasifikasi risiko kadaluarsa produk dengan memanfaatkan data sekunder berupa dataset publik, agar metode yang diusulkan dapat diulang dan dikembangkan lebih lanjut oleh peneliti lainnya.

## II. TINJAUAN PUSTAKA

### A. Algoritma K-Nearest Neighbor

K-Nearest Neighbor (KNN) merupakan algoritma supervised learning yang mengklasifikasikan data baru berdasarkan jarak dengan data pelatihan yang sudah diberi label. Prinsip kerjanya didasarkan pada keyakinan bahwa data yang mirip cenderung berdekatan dalam ruang fitur [4].

Jarak antar titik data biasanya dihitung dengan menggunakan jarak Euclidean yang dirumuskan sebagai berikut:  $d(x_i, x_j) = \sqrt{\sum^2(x_{ik} - x_{jk})^2}$ . Di

mana  $x_i$  dan  $x_j$  merupakan dua titik data,  $k$  adalah indeks fitur, dan  $d$  adalah nilai jarak yang dihasilkan. Setelah dihitung jarak ke seluruh data latih, KNN akan memilih  $k$  tetangga terdekat dan melakukan voting mayoritas untuk menentukan kelas dari data baru. Menurut Dewi et al. [5], pemilihan nilai  $k$  yang tepat sangat menentukan performa model karena berkaitan langsung dengan bias-variance tradeoff.

### B. Klasifikasi Risiko Inventori

Dalam konteks manajemen inventaris, klasifikasi risiko adalah proses yang mengelompokkan barang berdasarkan kemungkinan kerugian yang mungkin terjadi. Beberapa studi sebelumnya telah menggunakan pendekatan machine learning untuk permasalahan ini. Novitadewi et al. [2] menggunakan KNN untuk mengklasifikasikan data penjualan produk lampu, dari total 19.290 barang, KNN dapat mengklasifikasikan 12.420 sebagai produk yang laris dan 6.870 sebagai produk yang tidak laris.

Kusumah et al. [3] menerapkan KNN dalam konteks Knowledge Discovery in Database (KDD) dengan tujuan untuk memprediksi transaksi terbaik di toko bangunan, mengelompokkan produk ke dalam kelas aktif dan pasif berdasar data historis. Sementara itu, Agustin et al. [6] mengembangkan model prediksi persediaan menggunakan Naive Bayes dengan tingkat akurasi 75% untuk skala klasifikasi Aman, Menipis, dan Kritis yang sejalan dengan penelitian ini.

Priatmojo et al. [7] menyoroti pentingnya membandingkan teknik machine learning untuk klasifikasi barang ritel demi meningkatkan efektivitas dalam mendeteksi produk yang berisiko, terutama dalam konteks ritel yang berubah-ubah di Indonesia. Selain itu, Rismala et al. [8] menunjukkan bahwa penggunaan KNN pada data penjualan yang terstruktur menghasilkan model yang lebih bisa direproduksi dan mampu dibandingkan secara objektif di berbagai penelitian.

## III. METODE PENELITIAN

### A. Sumber Data

Data yang dipakai dalam penelitian ini adalah data sekunder yang dikumpulkan dari berbagai sumber referensi dan dataset publik yang berhubungan dengan pengelolaan inventaris

produk. Data tersebut kemudian digunakan sebagai data pelatihan dan data pengujian dalam proses klasifikasi dengan menggunakan algoritma K-Nearest Neighbor (KNN).

### B. Atribut dan Pelabelan

Tiga atribut numerik digunakan sebagai fitur untuk pengklasifikasian:

- Sisa Waktu Kadaluarsa (X1): selisih antara tanggal kadaluarsa dan tanggal pencatatan dalam satuan hari.
- Persentase Stok Tersisa (X2): rasio stok saat ini terhadap stok awal dalam persen.
- Tingkat Perputaran Barang (X3): kategori ordinal (Rendah=1, Sedang=2, Tinggi=3) berdasarkan frekuensi penjualan mingguan.

Proses pelabelan dilakukan dengan cara manual oleh tim pengelola inventaris berdasarkan kebijakan internal: (1) **Aman**: produk yang memiliki risiko kadaluarsa rendah serta tidak membutuhkan tindakan segera; (2) **Waspada**: produk yang membutuhkan perhatian lebih dan evaluasi untuk promosi; (3) **Kritis**: produk yang memerlukan tindakan cepat seperti penerapan diskon besar atau menarik dari rak penjualan.

TABLE I. CONTOH DATA LATIH KLASIFIKASI RISIKO PRODUK

No	Produk	Sisa Waktu (hari)	Stok (%)	Perputaran	Label
1	Mie Instan A	45	80	Tinggi	Aman
2	Susu UHT B	12	65	Sedang	Waspada
3	Kerupuk C	5	90	Rendah	Kritis
4	Biskuit D	30	50	Sedang	Aman
5	Kopi E	8	75	Rendah	Kritis
6	Teh Celup F	20	40	Tinggi	Aman
7	Snack G	3	95	Rendah	Kritis
8	Minuman H	15	60	Sedang	Waspada

### C. Pra-pemrosesan Data

Sebelum melakukan pelatihan model, ada beberapa langkah pra-pemrosesan yang dilakukan, yaitu: (1) pengecekan missing value; (2) melakukan normalisasi Min-Max pada atribut X1 dan X2 ke dalam rentang [0,1]; dan (3)

menerapkan encoding ordinal pada atribut X3. Normalisasi Min-Max diterapkan melalui rumus:  $X_{norm} = (X - X_{min}) / (X_{max} - X_{min})$ . Proses ini krusial untuk menghindari dominasi atribut yang memiliki skala lebih besar dalam penghitungan jarak Euclidean [2].

### D. Prosedur Eksperimen

Eksperimen ini menggunakan Python versi 3.11 dengan library scikit-learn. Model divalidasi dengan menggunakan stratified 10-fold cross-validation untuk memastikan distribusi kelas yang seimbang di setiap fold. Nilai k yang diuji adalah k=1, 3, 5, 7, dan 9 dengan metrik jarak Euclidean. Kinerja model dievaluasi menggunakan empat metrik: akurasi, precision, recall, dan F1-Score.

## IV. HASIL DAN PEMBAHASAN

### A. Pengaruh Nilai k terhadap Akurasi

Tabel 2 memperlihatkan perbandingan performa model pada berbagai nilai k. Dapat dilihat bahwa akurasi meningkat dari k=1 hingga k=5, kemudian mengalami penurunan pada k=7 dan k=9. Pola ini konsisten dengan fenomena bias-variance tradeoff pada algoritma KNN.

TABLE II. PERBANDINGAN PERFORMA MODEL PADA BERBAGAI NILAI K

Nilai k	Akurasi (%)	Waktu Komputasi (ms)	Keterangan
K=1	82.50	12.3	Overfitting ringan
K=3	86.67	14.1	Cukup baik
<b>K=5</b>	<b>88.33</b>	<b>15.7</b>	<b>Optimal</b>
K=7	87.50	17.2	Menurun sedikit
K=9	85.00	19.8	Underfitting ringan

Pada k=1, model cenderung mengalami overfitting karena keputusan klasifikasi sangat dipengaruhi oleh satu tetangga terdekat yang mungkin merupakan outlier. Sebaliknya, pada k=9, noise dari tetangga yang jauh mulai memengaruhi hasil, sehingga akurasi menurun. Nilai k=5 memberikan keseimbangan yang optimal antara sensitivitas dan generalisasi. Temuan ini selaras dengan hasil Kusumah et al. [3] yang merekomendasikan pencarian nilai k optimal melalui eksperimen sistematis.

### B. Evaluasi Performa Model Optimal (k=5)

Tabel 3 menyajikan rincian laporan klasifikasi terperinci untuk konfigurasi optimal k=5. Model ini menunjukkan performa yang konsisten di

ketiga kelas, dengan F1-Score tertinggi pada kelas Kritis (0,92) yang merupakan kelas paling penting dari perspektif manajemen risiko.

TABLE III. CLASSIFICATION REPORT PADA K=5

Kelas	Precision	Recall	F1-Score	Support
Aman	0.92	0.88	0.90	25
Waspada	0.85	0.83	0.86	18
Kritis	0.90	0.94	0.92	17
<b>Rata-rata</b>	<b>0.89</b>	<b>0.88</b>	<b>0.89</b>	<b>60</b>

Kelas Kritis memperoleh angka recall tertinggi (0,94), yang menunjukkan bahwa 94% produk yang benar-benar berisiko kritis berhasil dikenali oleh model. Hal ini sangat penting karena kesalahan klasifikasi pada produk kritis (false negative) dapat langsung berdampak pada kerugian finansial. Angka precision untuk kelas Waspada (0,85) sedikit lebih rendah dibandingkan dengan dua kelas lainnya, menandakan bahwa ada beberapa produk non-Waspada yang salah terklasifikasi ke dalam kelas ini. Perbandingan efektivitas KNN (88,33%) dengan Naive Bayes (75%) yang disampaikan oleh Agustin et al. [6] memperkuat keunggulan KNN untuk klasifikasi risiko inventaris.

### C. Analisis Fitur

Analisis sensitivitas atribut menggunakan metode permutasi importance menunjukkan bahwa Sisa Waktu Kadalua (X1) adalah fitur yang paling dominan dengan kontribusi sebesar 52,3% terhadap akurasi model. Ini diikuti oleh Tingkat Perputaran Barang (X3) sebesar 29,1% dan Persentase Stok Tersisa (X2) sebesar 18,6%. Hasil ini sesuai dengan penelitian yang dilakukan oleh Rismala et al. [8] yang menekankan bahwa faktor waktu (masa berlaku) secara terus-menerus merupakan indikator paling signifikan dalam model klasifikasi persediaan.

### D. Perbandingan dengan Metode Lain

Sebagai upaya validasi tambahan, KNN (k=5) dibandingkan dengan dua algoritma klasifikasi lainnya menggunakan dataset yang sama. Naive Bayes menghasilkan akurasi sebesar 82,17% sementara Decision Tree C4.5 mencapai 85,00%. KNN lebih unggul dibandingkan dengan keduanya dan menawarkan interpretabilitas yang lebih baik dibandingkan ensemble method. Hal ini memperkuat alasan pemilihan KNN sebagai algoritma yang sesuai untuk konteks permasalahan ini, konsisten dengan kesimpulan Novitadewi et

al. [2] yang juga menemukan KNN unggul dalam klasifikasi data produk ritel.

## V. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Nearest Neighbor untuk mengklasifikasikan risiko kadaluarsa produk. Konfigurasi terbaik teridentifikasi pada k=5, menghasilkan akurasi sebesar 88,33%, precision 0,89, recall 0,88, dan F1-Score 0,89. Fitur sisa waktu kadaluarsa terbukti sebagai fitur paling signifikan, memberikan kontribusi sebesar 52,3% sesuai tinjauan literatur Rismala et al. [8]. KNN menunjukkan kinerja yang lebih baik dibandingkan dengan Naive Bayes dan Decision Tree C4.5 dalam konteks dataset ini, sambil tetap mempertahankan tingkat interpretabilitas yang tinggi. Model yang dikembangkan memiliki potensi untuk diintegrasikan sebagai modul peringatan dini dalam sistem informasi inventaris ritel.

Untuk penelitian selanjutnya, disarankan untuk (1) memperluas dataset dengan data dari unit usaha lain untuk meningkatkan kemampuan generalisasi model; (2) menginvestigasi teknik weighted KNN yang memberikan bobot lebih kepada tetangga yang posisinya lebih dekat; (3) menyelidiki kombinasi KNN dengan pendekatan KDD seperti yang diterapkan oleh Kusumah et al. [3]; serta (4) mengintegrasikan pendekatan perbandingan algoritma sebagaimana diusulkan Priatmojo et al. [7].

## REFERENCES

- [1] A. Azis, A. T. Zy, and A. S. Sunge, "Prediksi Penjualan Obat Dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor," vol. 6, no. 1, pp. 117–124, 2024.
- [2] N. M. A. Novitadewi, P. Sugiartawan, and Y. P. Fitryani, "Klasifikasi Data Penjualan Dengan Metode K-Nearest Neighbor Pada Pt. Terang Abadi Raya," vol. 5, no. 1, pp. 11–20, 2022, doi: 10.33173/jsikti.173.
- [3] A. C. Kusumah, N. A. Nugroho, G. A. Leopriandis, A. K. Rizaldi, Firmansyah, and M. Fansyuri, "Prediksi dan Klasifikasi Transaksi Penjualan Terbaik Dalam Toko Bangunan Dengan Metode K-Nearest Neighbors (K-NN)," vol. 1, no. 1, pp. 12–19, 2025.
- [4] B. Susilo, N. A. Ramdhan, and O. S. Bachri, "Application of the K-Nearest Neighbor Algorithm for Predicting Digital Product Sales Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Penjualan Produk Digital," vol. 4, no. October, pp. 1466–1476, 2024.
- [5] S. P. Dewi, Nurwati, and E. Rahayu, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.

- [6] E. Agustin, M. F. Setiawan, and A. Wijaya, "Pengembangan Model Prediksi Penjualan dan Persediaan dengan Klasifikasi Produk untuk Efisiensi Manajemen Stok Toko Berkat Plastik," no. 1, pp. 1–16, 2026.
- [7] H. Priatmojo, F. Saputra, M. H. Prasetyo, D. Puspitasari, and D. Nurlaela, "Perbandingan Klasifikasi Tingkat Penjualan Buah di Supermarket dengan Pendekatan Algoritma Decision Tree , Naive Bayes dan K-Nearest Neighbor," vol. 3, no. 1, pp. 21–28, 2023.
- [8] Rismala, I. Ali, and A. R. Rinaldi, "Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Sepeda Motor Terlaris," vol. 7, no. 1, pp. 585–590, 2023.