

## Penerapan Metode K-Nearest Neighbor untuk Identifikasi Produk Nike Paling Laris Terjual

Venia Restreva Danestiara<sup>1)</sup>, M. Achya Arifudin<sup>2)</sup>, Aura Tifa Aprilla<sup>3)</sup>, Dani Setiawan<sup>4)</sup>

<sup>1)</sup> Informatika, Universitas Informatika dan Bisnis Indonesia

<sup>2)</sup> Sistem Informasi, Universitas Informatika dan Bisnis Indonesia

Email: veniarestreva@unibi.ac.id; achyaarifudin@unibi.ac.id; auratifa14@gmail.com;  
daniset240503@gmail.com

### Abstrak

Industri ritel menghadapi persaingan yang semakin ketat, sehingga kemampuan untuk memprediksi produk terlaris menjadi faktor penting dalam memaksimalkan keuntungan. Nike, sebagai salah satu produsen sepatu olahraga terkemuka, mengalami penurunan signifikan selama pandemi COVID-19 dengan pendapatan turun sebesar 38% dan kerugian mencapai \$790 juta pada tahun 2020. Oleh karena itu, prediksi akurat terhadap produk terlaris periode 2020-2021 menjadi krusial untuk memulihkan kinerja pasar global. Penelitian ini menerapkan algoritma K-Nearest Neighbors (KNN) untuk memprediksi produk Nike yang paling laris menggunakan dataset *open-source* dari Kaggle. Penelitian ini melibatkan tahapan *preprocessing* data, seperti *feature selection*, normalisasi, dan label *encoding*. Model KNN diimplementasikan dengan variasi nilai  $k$  untuk mengevaluasi pengaruhnya terhadap akurasi. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, *recall*, dan F1-Score. Hasil penelitian menunjukkan akurasi model mencapai 87% dengan keseimbangan baik antara *precision* (0.89) dan *recall* (0.87). "Women's Street Footwear" teridentifikasi sebagai produk terlaris dengan 57,737 unit terjual. Model menunjukkan performa optimal dengan akurasi 75%. Penelitian ini memberikan kontribusi signifikan dalam pengambilan keputusan strategis untuk optimasi inventory dan strategi penjualan Nike.

**Kata Kunci:** K-Nearest Neighbors, Prediksi Penjualan, Nike, Produk Terlaris, Akurasi.

### Abstract

*The retail industry is facing increasing competition, so the ability to predict best-selling products is an important factor in maximizing profits. Nike, as one of the leading sports shoe manufacturers, experienced a significant decline during the COVID-19 pandemic with revenues falling by 38% and losses reaching \$790 million in 2020. Therefore, accurate prediction of the best-selling products for 2020-2021 is crucial to recover the global market performance. This research applies the K-Nearest Neighbors (KNN) algorithm to predict Nike's best-selling products using an open-source dataset from Kaggle. This research involves data preprocessing stages, such as feature selection, normalization, and label encoding. The KNN model was implemented with varying  $k$  values to evaluate its effect on accuracy. Model evaluation is done using accuracy, precision, recall, and F1-Score metrics. The results showed that the model accuracy reached 87% with a good balance between precision (0.89) and recall (0.87). "Women's Street Footwear" was identified as the best-selling product with 57,737 units sold. The model showed optimal performance with 75% accuracy. This research makes a significant contribution to strategic decision making for Nike's inventory optimization and sales strategy.*

**Keywords:** K-Nearest Neighbors, Sales Prediction, Nike, Best-Selling Product, Accuracy.

## 1 PENDAHULUAN

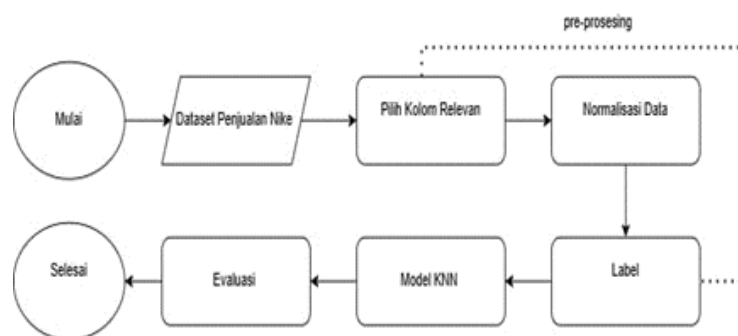
Dalam industri ritel yang semakin kompetitif, kemampuan untuk memprediksi produk yang akan laris terjual menjadi faktor kunci dalam memaksimalkan keuntungan perusahaan. Salah satu produsen sepatu olahraga terkemuka di dunia ialah Nike. Nike termasuk bisnis yang mendominasi industri penjualan sepatu atletik (Okhtavia dan Setiawan). Prediksi akurat produk terlaris pada periode 2020-2021 sangat krusial untuk meningkatkan keuntungan dan efisiensi operasional. Nike mengalami penurunan yang signifikan dan menjadi krusial bagi kelangsungan bisnisnya (Zameo, 2024). Penutupan sekitar 50% dari toko fisik di wilayah seperti China, Macao, dan Hong Kong akibat pandemi COVID-19 menyebabkan kerugian besar, dengan Nike kehilangan sekitar \$790 juta dan pendapatan turun 38% pada kuartal terakhir tahun fiskal 2020 (Nike, 2021). Oleh karena itu, penting bagi Nike untuk dapat memprediksi produk yang akan memiliki permintaan tinggi guna meraih kembali kinerja yang optimal di pasar global.

Penelitian ini mengajukan implementasi algoritma K-Nearest Neighbor (KNN) untuk mengidentifikasi pola produk Nike yang paling laris terjual berdasarkan dataset dari Kaggle.com yang bersifat open source. Metode K-Nearest Neighbor (K-NN) dipilih karena kemampuannya dalam melakukan klasifikasi berdasarkan kedekatan jarak suatu data dengan data lainnya, dimana meskipun tergolong sederhana dalam implementasi, metode ini memiliki tingkat akurasi yang tinggi (Yolanda dan Fahmi, 2020). Dengan demikian, algoritma ini dapat menjadi alat yang berguna dalam pengambilan keputusan bisnis untuk meningkatkan strategi penjualan.

Salah satu penelitian terdahulu menggunakan algoritma K-Nearest Neighbor (KNN) untuk memprediksi produk terlaris dengan hasil yang signifikan, berhasil memprediksi 6 produk elektronik terlaris dari 22 produk yang terjual, yaitu CTV, Lemari Es, DVD, Speaker, Mesin Cuci, dan LCD, dengan akurasi klasifikasi mencapai 92,51% (Amalia, 2024). Penelitian terdahulu lainnya menggunakan KNN untuk memprediksi penjualan produk meubel terlaris, di mana lemari pakaian menjadi produk terlaris dengan 90 unit terjual, dan akurasi klasifikasi penjualan mencapai 87,51%. Kedua penelitian ini membuktikan kemampuan KNN dalam memberikan prediksi yang akurat untuk berbagai kategori produk (Suryadi dkk, 2022).

## 2 METODE PENELITIAN

Berikut beberapa tahapan yang dilakukan dalam penelitian ini:



Gambar 1. Alur penelitian

Pada Gambar 1, menunjukkan alur penelitian yang menggunakan pendekatan sistematis dalam pengolahan data dan pemodelan KNN yang terdiri dari dua jalur utama proses. Jalur pertama berfokus pada *pre-processing* data yang dimulai dari tahap pengumpulan dataset perputaran Nike, dilanjutkan dengan seleksi kolom-kolom yang relevan, dan normalisasi data untuk memastikan konsistensi skala data. Sementara itu, jalur kedua mencakup proses seleksi dan evaluasi yang menghasilkan model KNN untuk mengklasifikasikan data menjadi label-label yang telah ditentukan.

Kedua jalur ini terintegrasi dalam satu sistem dimana hasil *pre-processing* data dari jalur pertama menjadi input untuk proses pemodelan pada jalur kedua, memastikan kualitas dan keakuratan hasil klasifikasi yang dihasilkan oleh model KNN.

## 2.1 Dataset

	Invoice Date	Product	Region	Retailer	Sales Method	State	Price per unit	Total sales	Units sold
0	01-01-2020	Men's Street Footwear	Northeast	Foot Locker	In-store	New York	50	6000	120
1	02-01-2020	Men's Athletic Footwear	Northeast	Foot Locker	In-store	New York	50	5000	100
2	03-01-2020	Women's Street Footwear	Northeast	Foot Locker	In-store	New York	40	4000	100
3	04-01-2020	Women's Athletic Footwear	Northeast	Foot Locker	In-store	New York	45	3825	85
4	05-01-2020	Men's Apparel	Northeast	Foot Locker	In-store	New York	60	5400	90
--	--	--	--	--	--	--	--	--	--
9355	06-07-2021	Women's Athletic Footwear	West	West Gear	Outlet	Idaho	38	60	16
9356	05-04-2021	Women's Street Footwear	West	West Gear	Outlet	Idaho	19	31	16
9357	05-05-2021	Women's Street Footwear	West	West Gear	Outlet	Idaho	18	33	18
9358	04-06-2021	Women's Street Footwear	West	West Gear	Outlet	Idaho	34	63	19
9359	06-07-2021	Women's Street Footwear	West	West Gear	Outlet	Idaho	34	95	28

Gambar 2. Dataset

Dataset yang ditunjukkan pada Gambar 2. diambil dari situs Kaggle, sebuah platform yang menyediakan berbagai dataset open-source yang dapat digunakan untuk analisis data. Dataset ini berisi 9.359 data mencakup informasi penjualan produk Nike di Amerika Serikat pada periode 2020-2021, termasuk data transaksi, jenis produk, harga, jumlah unit terjual, wilayah, harga per-unit dan total penjualan.

## 2.2 Preprocessing Data

*Pre-processing* data digunakan untuk mengambil informasi dari suatu data yang masih mentah. Dan menghilangkan *noise* yang terdapat dalam data mentah, sehingga dapat diproses selanjutnya (Kasanah dkk, 2019). *Pre-processing* data dalam penelitian ini meliputi pemilihan kolom relevan (*feature selection*) untuk memastikan fitur yang digunakan berkontribusi pada analisis, normalisasi data untuk menyamakan skala antar fitur menggunakan Min-Max atau Z-score Normalization, dan konversi label kategori menjadi numerik melalui Label *Encoding* agar dapat diproses oleh metode KNN. Selain itu, data yang besar dapat diperkecil dengan teknik seperti agregasi data *cube*, pemilihan atribut yang relevan, dan pengurangan dimensi untuk membuat penyimpanan dan analisis lebih efisien (Hans, 2024).

## 2.3 Metode K-Nearest Neighbor

K-Nearest Neighbor (KNN) adalah algoritma yang mengklasifikasikan objek baru berdasarkan mayoritas kelas dari tetangga terdekatnya. KNN termasuk algoritma supervised learning, yang mana hasil dari query instance baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN (Dewi dkk, 2022). Tujuan dari algoritma KNN adalah untuk untuk mengklasifikasikan objek baru berdasarkan atribut dan training samples, dan hasil dari sampel uji yang baru diklasifikasikan berdasarkan sebagian besar kategori yang ada pada KNN (Liantoni, 2015). Dalam algoritma KNN, nilai *neighbor* (K) yang terbaik bergantung pada jumlah data latih. Nilai *neighbor* (K) yang terbaik tidak diukur dari besar atau kecilnya nilai *k* (Herman, 2020).

Tahapan-tahapan untuk menghitung algoritma KNN yaitu, menentukan nilai  $k$ , menghitung kuadrat jarak *euclid* (*query instance*) masing-masing objek terhadap data latih (*training*), mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *euclid* terkecil, mengumpulkan label kelas  $Y$  (Klasifikasi K-Nearest Neighbor). Berikut rumus untuk menghitung algoritma KNN dapat dilihat pada Tabel 1.

Tabel 1. Rumus KNN

Rumus	Keterangan
$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$	<ul style="list-style-type: none"> <li>▪ <math>p_i</math> = nilai fitur pada data <i>training</i>.</li> <li>▪ <math>q_i</math> = nilai fitur pada data <i>testing</i>.</li> <li>▪ <math>n</math> = jumlah fitur (dimensi data).</li> </ul>

### 2.4 Evaluasi

Tahap evaluasi model ini dilakukan untuk mengetahui kinerja suatu algoritma dalam mengetahui keakuratan dari model diperlukan perhitungan confusion matrix (Silalahi dan Simanullang, 2023).

Tabel 2. Confusion Matrix

Confusion Matrix	Nilai Aktual		
	Positif	Negatif	
Nilai Prediksi	Positif	True Positives	False Positive
	Negatif	False Negatives	True Negatives

Perhatikan Tabel 2, confusion matrix digunakan untuk membandingkan prediksi dan data aktual, serta menghitung akurasi, presisi, recall, dan F1-score (Said dkk, 2022). Rumus untuk masing-masing pengukuran ini dapat dilihat pada Tabel 3 berikut:

Tabel 3. Pengukuran Nilai Accuracy, Precision, Recall dan F1-Score

Pengukuran	Rumus
Accuracy	$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$
Precision	$Precision = \frac{TP}{TP + FP}$
Recall	$Recall = \frac{TP}{TP + FN}$
F1-Score	$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision}$

### 3 HASIL DAN PEMBAHASAN

#### 3.1 Label Encoding

Hasil encoding 'Product', 'Region', 'Sales Method', 'State':

	Product	Region	Sales Method	State	Units Sold
0	2	1	0	29	120
1	1	1	0	29	100
2	5	1	0	29	100
3	4	1	0	29	85
4	0	1	0	29	90

Gambar 3. Hasil Encoding

Pada tahap *pre-processing* data, dilakukan *encoding* terhadap beberapa fitur kategorikal dalam dataset. Kolom *Product*, *Region*, *Sales Method*, dan *State* yang sebelumnya berupa data kategorikal diubah menjadi format numerik. Setiap kategori dalam kolom-kolom ini diberi nilai numerik untuk mempermudah pemrosesan lebih lanjut oleh model *machine learning*. Perhatikan pada Gambar 3, kolom *Product* merepresentasikan jenis produk yang telah diberi label numerik, seperti 0, 1, 2, dan seterusnya. Kolom *Region* menunjukkan wilayah penjualan, yang juga diubah menjadi angka. Kolom *Sales Method* merepresentasikan metode penjualan, dengan nilai numerik untuk kategori tertentu. Kolom *State* menunjukkan status wilayah atau area tertentu, dengan angka sebagai pengganti nama area. Kolom *Units Sold* tetap dalam format numerik asli karena sudah berbentuk kuantitatif dan merepresentasikan jumlah unit yang terjual. Proses ini memastikan bahwa data dapat diproses secara efisien oleh algoritma pembelajaran mesin yang membutuhkan input numerik, sekaligus mempertahankan informasi penting dari dataset asli.

#### 3.2 Pengelompokkan Data

Tampilan Produk, Units Sold, dan Jumlah Produk:

	Product	Jumlah Produk	Units Sold
0	0	1560	30085
1	1	1560	42429
2	2	1560	57737
3	3	1560	42356
4	4	1560	31068
5	5	1560	38309

Gambar 4. Pengelompokkan Data

Hasil analisis penjualan produk, data dikelompokkan berdasarkan kategori produk, jumlah produk, dan *units sold* (jumlah unit yang terjual). Berdasarkan gambar di bawah, terdapat enam kategori produk yang masing-masing memiliki jumlah produk sebesar 1.560 unit. Namun, terdapat perbedaan signifikan pada total units sold untuk masing-masing kategori produk, yang menunjukkan variasi tingkat penjualan.

### 3.3 Clustering Produk Terlaris dan Tidak Laris

	Product	Units Sold	Klasifikasi
0	0	30085	Tidak Laris
1	1	42429	Terlaris
2	2	57737	Terlaris
3	3	42356	Terlaris
4	4	31068	Tidak Laris
5	5	38309	Tidak Laris

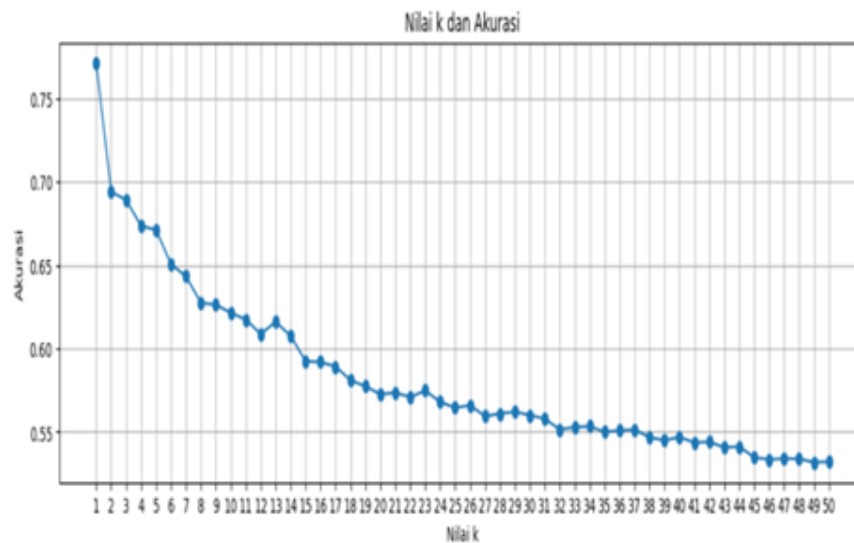
Jumlah Produk Terlaris: 3

Jumlah Produk Tidak Laris: 3

Gambar 5. Clustering Produk

Pada Gambar 5 *clustering* produk, ditampilkan hasil klasifikasi penjualan produk berdasarkan jumlah *Units Sold*. Dari klasifikasi tersebut, enam kategori produk dibandingkan untuk menentukan mana yang termasuk Terlaris dan Tidak Laris. Hasilnya menunjukkan bahwa tiga kategori produk berhasil diklasifikasikan sebagai Terlaris, sedangkan tiga lainnya termasuk dalam kategori Tidak Laris. Produk dengan jumlah *Units Sold* tertinggi mencapai 57.737 unit, sementara produk dengan penjualan terendah hanya mencapai 30.085 unit. Meskipun setiap kategori memiliki jumlah produk yang sama, yaitu 1.560 unit, terdapat perbedaan signifikan dalam performa penjualannya. Berdasarkan hasil yang ditampilkan, Produk yang Terlaris ditunjukkan oleh Produk 2 dan Units Sold 57,737 dengan nama produk yang dimaksud adalah “*Women’s Street Footwear*”.

### 3.4 Penerapan Model K-Nearest Neighbor



Gambar 6. Grafik Nilai K dan Akurasi

Berdasarkan hasil pengujian, grafik pada Gambar 6 menunjukkan hubungan antara nilai  $k$  (jumlah tetangga terdekat) dan akurasi model K-Nearest Neighbors (KNN) dalam klasifikasi. Gambar tersebut menunjukkan hubungan antara nilai  $k$  (jumlah tetangga terdekat) dengan akurasi model K-Nearest Neighbors (KNN) dalam memprediksi produk terbaik. Grafik memperlihatkan bahwa akurasi tertinggi tercapai pada nilai  $k = 1$ , dengan akurasi sekitar 75%. Hal ini menunjukkan bahwa mempertimbangkan hanya tetangga terdekat memberikan hasil yang paling akurat pada dataset ini. Namun, seiring dengan meningkatnya nilai  $k$ , akurasi model cenderung menurun. Penurunan ini disebabkan oleh pengaruh tetangga yang kurang relevan ketika jumlah tetangga yang dimasukkan dalam perhitungan semakin besar. Setelah nilai  $k = 20$ , akurasi cenderung stabil di kisaran 55-60%, dengan fluktuasi yang kecil meskipun nilai  $k$  terus meningkat. Hal ini mengindikasikan bahwa model mulai kehilangan sensitivitas terhadap data spesifik.

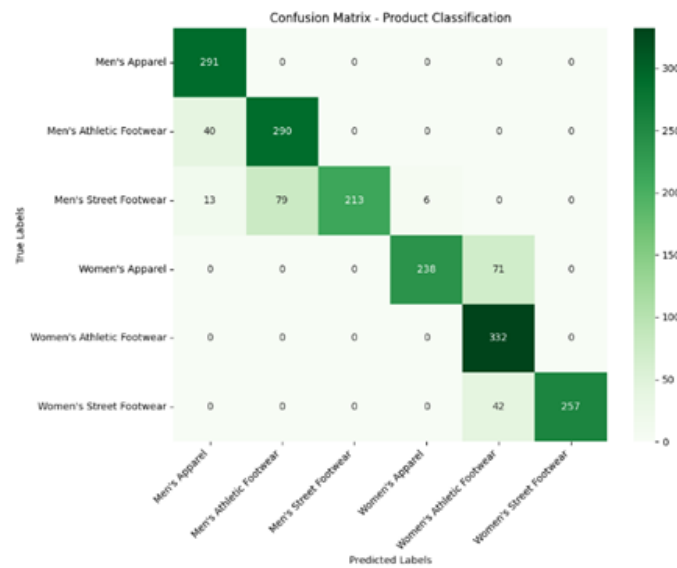
### 3.5 Evaluasi Model

Berdasarkan hasil pengujian model klasifikasi, diperoleh kinerja yang cukup baik dengan akurasi keseluruhan sebesar 0.87 atau 87%. Model menunjukkan keseimbangan yang baik antara precision dan recall, dengan nilai macro average precision 0.89 dan recall 0.87, menghasilkan F1-Score yang konsisten sebesar 0.87. Evaluasi kinerja model secara keseluruhan dapat dilihat pada Tabel 4 berikut.

Tabel 4. Evaluasi Kinerja Model

Metrik Evaluasi	Nilai
Akurasi	0.87
Precisiom	0.89
Recall	0.87
F1-Score	0.87

### 3.6 Confusion Matrix



Gambar 7. Confusion Matrix

Confusion matrix pada Gambar 7 menunjukkan performa model dalam mengklasifikasikan enam kategori produk: *Men's Apparel*, *Men's Athletic Footwear*, *Men's Street Footwear*, *Women's Apparel*, *Women's Athletic Footwear*, dan *Women's Street Footwear*. Model menunjukkan performa yang baik dengan nilai tinggi pada diagonal utama, menandakan jumlah prediksi yang benar di masing-masing kategori. Misalnya, kategori *Men's Apparel* memiliki 291 prediksi benar, *Men's Athletic Footwear* 290, *Men's Street Footwear* 213, *Women's Apparel* 238, *Women's Athletic Footwear* 332, dan *Women's Street Footwear* 257. Namun, terdapat beberapa kesalahan klasifikasi, seperti 40 sampel *Men's Athletic Footwear* yang salah diprediksi menjadi *Men's Street Footwear*, serta 79 sampel *Men's Street Footwear* yang salah diklasifikasikan sebagai *Men's Athletic Footwear*. Kesalahan serupa juga terjadi pada *Women's Apparel* dengan 71 sampel yang salah diprediksi sebagai *Women's Street Footwear*. Meski begitu, model secara keseluruhan menunjukkan kemampuan yang baik dalam membedakan kategori produk, meskipun terdapat peluang untuk perbaikan lebih lanjut dalam mengurangi kesalahan prediksi antar kategori tertentu.

#### 4 SIMPULAN

Berdasarkan hasil penelitian implementasi algoritma K-Nearest Neighbor (KNN) untuk memprediksi produk Nike terlaris, dapat disimpulkan bahwa model yang dikembangkan menunjukkan performa yang sangat memuaskan dengan akurasi keseluruhan mencapai 87%. Model ini menunjukkan keseimbangan yang baik antara *precision* (0.89) dan *recall* (0.87), dengan *F1-Score* yang konsisten sebesar 0.87, mengindikasikan stabilitas dan kehandalan model dalam melakukan prediksi. Analisis data mengungkapkan bahwa "*Women's Street Footwear*" (*Product 2*) menjadi produk terlaris dengan pencapaian penjualan sebesar 57,737 unit, memberikan insight strategis yang berharga untuk manajemen *inventory* dan strategi pemasaran ke depan. Dalam pengujian parameter, ditemukan bahwa nilai  $k=1$  memberikan hasil optimal dengan akurasi mencapai 75%, menunjukkan bahwa pendekatan tetangga terdekat tunggal paling efektif untuk karakteristik dataset ini. Hasil penelitian ini membuktikan bahwa algoritma KNN dapat menjadi alat yang efektif dalam memprediksi tren penjualan produk Nike, yang dapat membantu perusahaan dalam mengambil keputusan strategis untuk mengoptimalkan *inventory* dan meningkatkan performa penjualan di masa mendatang. Rekomendasi penelitian selanjutnya, disarankan untuk mengeksplorasi algoritma lain, seperti *Random Forest* atau *Neural Networks*, guna membandingkan performa dan akurasi prediksi. Penelitian juga dapat menambahkan variabel data, seperti ulasan pelanggan atau tren media sosial,

untuk meningkatkan kualitas prediksi. Selain itu, penerapan model pada skala global perlu dilakukan untuk memahami perbedaan pola penjualan di berbagai wilayah. Optimalisasi parameter algoritma KNN dan integrasi data *real-time* juga direkomendasikan agar prediksi lebih dinamis dan relevan dengan kondisi pasar terkini.

#### DAFTAR PUSTAKA

- R. Okhtavia And M. B. Setiawan, “Pengaruh Kualitas Produk, Citra Merek, Dan Inovasi Produk Terhadap Keputusan Pembelian Konsumen Sepatu Olahraga Nike (Studi Pada Pengguna Sepatu Olahraga Nike Di Kota Kendal)”.
- S. Zameo, “Sports Brands And Covid-19: A Nike Case Study | Textmaster,” The International Expansion Blog. Accessed: Nov. 20, 2024. [Online]. Available: <https://www.textmaster.com/blog/sports-brands-and-covid-19-a-nike-case-study/>
- “Nike, Inc. - Investor Relations - Nike, Inc. Reports Fiscal 2021 Second Quarter Results.” Accessed: Nov. 20, 2024. [Online]. Available: <https://investors.nike.com/investors/news-events-and-reports/investor-news/investor-news-details/2020/nike-inc.-reports-fiscal-2021-second-quarter-results/default.aspx>
- I. Yolanda And H. Fahmi, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada Pt.Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor,” J. Ilmu Komput. Dan Sist. Inf. Jikoms, Vol. 3, No. 1.1, Pp. 9–15, 2020, Doi: 10.9767/Jikoms.V3i1.1.83.
- N. 13540268 Yulia Rizki Amalia, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode K-Nearest Neighbor (Studi Kasus : Pt. Bintang Multi Sarana Palembang),” Diploma, Uin Raden Fatah Palembang, 2018. Accessed: Nov. 17, 2024. [Online]. Available: <http://perpustakaan.ac.id>
- L. Suryadi, N. Ngajiyanto, N. E. Pratiwi, F. Ardhy, And P. Riswanto, “Penerapan Data Mining Prediksi Penjualan Mebel Terlaris Menggunakan Metode K-Nearest Neighbor(K-Nn) (Studi Kasus : Toko Zerita Meubel),” Jusim J. Sist. Inf. Musirawas, Vol. 7, No. 2, Pp. 174–184, Dec. 2022, Doi: 10.32767/Jusim.V7i2.1697.
- A. N. Kasanah, M. Muladi, And U. Pujiyanto, “Penerapan Teknik Smote Untuk Mengatasi Imbalance Class Dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma Knn,” J. Resti Rekayasa Sist. Dan Teknol. Inf., Vol. 3, No. 2, Art. No. 2, Aug. 2019, Doi: 10.29207/Resti.V3i2.945.
- R. Hans, “Pentingnya Preprocessing Dalam Pengolahan Data Statistik.” Accessed: Nov. 18, 2024. [Online]. Available: <https://dqlab.id/pentingnya-preprocessing-dalam-pengolahan-data-statistik>
- S. P. Dewi, N. Nurwati, And E. Rahayu, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor,” Build. Inform. Technol. Sci. Bits, Vol. 3, No. 4, Art. No. 4, Mar. 2022, Doi: 10.47065/Bits.V3i4.1408.
- F. Liantoni, “Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor,” Ultim. J. Tek. Inform., Vol. 7, No. 2, Pp. 98–104, 2015, Doi: 10.31937/Ti.V7i2.356.
- I. H. Herman, “Penggunaan K-Nearest Neighbor (Knn) Untuk Mengidentifikasi Citra Batik Pewarna Alami Dan Pewarna Sintetis Berdasarkan Warna,” Pros. Semin. Nas. Mhs. Bid. Ilmu Komput. Dan Apl., Vol. 1, No. 2, Art. No. 2, Nov. 2020, Accessed: Nov. 20, 2024. [Online]. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/596>
- Universitas Methodist Indonesia, A. P. Silalahi, And H. G. Simanullang, “Supervised Learning Metode K-Nearest Neighbor Untuk Prediksi Diabetes Pada Wanita,” Methomika J. Manaj. Inform. Dan Komputerisasi Akunt., Vol. 7, No. 1, Pp. 144–149, Apr. 2023, Doi: 10.46880/Jmika.Vol7no1.Pp144-149.

H. Said, N. H. Matondang, And H. N. Irmanda, “Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi,” *Techno.Com*, Vol. 21, No. 2, Pp. 256–267, May 2022, Doi: 10.33633/Tc.V21i2.5901.