Buletin Ilmiah
Ilmu Komputer dan Multimedia

BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 3, No. 4, November 2025
ISSN 3024-8248 (media online)
Hal 550-557

Analisis Penjualan Nike dengan *Random Forest* dan *K-Means* untuk Identifikasi Pola dan Produk Terlaris

Nurya Herlina Sari[1], Noor Latifah[2]

[12]Fakultas Teknik, Program Studi Sistem Informasi, Universitas Muria Kudus, Kudus, Indonesia
Email: [1]202253064@std.umk.ac.id, [2]noor.latifah@umk.ac.id

Abstrak−Penelitian ini bertujuan untuk menganalisis pola penjualan produk Nike dengan memanfaatkan algoritma *Random Forest* dan *K-Means Clustering*. Dataset diambil dari Kaggle dan diproses dengan *Google Colab* untuk mengenali faktor-faktor yang memengaruhi volume penjualan serta segmentasi produk berdasarkan kesamaan karakteristik. *Random Forest* diterapkan untuk mengidentifikasi variabel yang paling berpengaruh terhadap penjualan, sementara *K-Means* digunakan untuk mengelompokkan produk ke dalam berbagai *Cluster*. Visualisasi menunjukkan bahwa harga adalah faktor utama yang paling memengaruhi penjualan, diikuti dengan penilaian pelanggan. Dengan menggunakan *Elbow Method*, ditemukan jumlah *Cluster* optimal sebanyak tiga, yaitu produk dengan harga rendah dan penjualan menengah, produk premium dengan penjualan rendah, serta produk dengan harga menengah dan penjualan tinggi. Segmentasi ini memberikan wawasan yang jelas tentang pola penjualan dan kelompok produk yang berpotensi besar untuk diperluas. Temuan dari penelitian ini diharapkan menjadi pedoman dalam strategi pemasaran, manajemen persediaan, dan inovasi produk berdasarkan atribut penjualan yang paling berpengaruh.

Kata Kunci: Data Mining; *K-Means*; *Random Forest*; Penjualan; Nike

Abstract−*This study aims to analyze Nike product sales patterns using the Random Forest and K-Means Clustering algorithms. The dataset was taken from Kaggle and processed with Google Colab to identify factors that influence sales volume and product segmentation based on characteristic similarities. Random Forest was applied to identify the variables that most influence sales, while K-Means was used to group products into various Clusters. Visualizations show that price is the main factor that most influences sales, followed by customer ratings. Using the Elbow Method , the optimal number of Clusters was found to be three: low-priced and mid-selling products, premium products with low sales, and mid-priced and high-selling products. This segmentation provides clear insights into sales patterns and product groups with great potential for expansion. The findings of this study are expected to guide marketing strategies, inventory management, and product innovation based on the most influential sales attributes.*

Keywords: *Data Mining; K-Means; Random Forest; Sale; Nike*

1. PENDAHULUAN

Kemajuan teknologi digital telah memberikan pengaruh besar pada berbagai sektor industri, termasuk sektor ritel dan olahraga. Perusahaan-perusahaan besar seperti Nike sekarang menggunakan data penjualan untuk menganalisis perilaku konsumen dan memperbaiki strategi pemasaran. Kemampuan dalam mengumpulkan, menganalisis, dan memanfaatkan data dengan efektif adalah faktor penting bagi perusahaan untuk memahami pola pasar, perilaku konsumen, serta tren penjualan (Ramadhan et al., 2025). Nike, sebagai salah satu merek internasional di dalam sektor olahraga, menawarkan berbagai macam produk dengan karakteristik yang bervariasi, mulai dari sepatu untuk berlari, sepatu untuk basket, pakaian olahraga, hingga aksesori pendukung lainnya. Setiap produk memiliki perbedaan dalam harga, desain, ulasan pelanggan, mutu, serta tingkat penjualannya. Pola penjualan yang tidak selalu serupa antara berbagai produk disebabkan oleh perbedaan karakteristik ini. Oleh sebab itu, analisis penting dilakukan untuk dapat memahami faktor-faktor yang berpengaruh terhadap penjualan dan bagaimana produk bisa dikelompokkan berdasarkan karakteristik tertentu.

Pemanfaatan teknik machine learning dalam analisis penjualan dapat memberikan keuntungan tambahan karena dapat memproses data dalam volume besar dan menciptakan pola secara otomatis. Dua pendekatan yang biasa diterapkan dalam analisis penjualan adalah *Random Forest* dan *K-Means Clustering*. penggunaan algoritma *Random Forest*, yang sudah terbukti efektif dalam menangani data yang kompleks dan dapat memberikan hasil prediksi yang andal (Sari et al., 2025). *Random Forest* merupakan algoritma *machine learning supervised* yang sudah terbukti paling efektif dalam klasifikasi teks. Metode ensemble ini menurunkan kemungkinan overfitting dan memperbaiki kemampuan generalisasi model, menjadikan *Random Forest* sebagai pilihan yang

BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 3, No. 4, November 2025
ISSN 3024-8248 (media online)
Hal 550-557

ideal untuk tugas klasifikasi (Artikel & Info, 2025). Algoritma ini memanfaatkan sejumlah kecil atribut dan subset data secara acak untuk melatih setiap pohon, lalu menggabungkan hasil prediksi untuk menetapkan klasifikasi akhir (Alfajr & Defiyanti, 2024). Dengan menganalisis pentingnya fitur, perusahaan dapat mengidentifikasi elemen yang paling berpengaruh terhadap keberhasilan penjualan produk tertentu. Data ini bisa dijadikan pedoman dalam proses pengambilan keputusan seperti menentukan harga, meningkatkan kualitas produk, atau mengembangkan fitur-fitur baru pada produk tertentu. Metode ini dapat menyajikan sumber informasi tentang tingkat pentingnya setiap variabel, sehingga dapat diidentifikasi variabel mana yang memiliki pengaruh terbesar. *K-Means* juga dapat digunakan untuk mengategorikan produk ke dalam beberapa kluster berdasarkan persamaan karakteristik. Selain itu, algoritma *K-Means Clustering* telah berhasil digunakan untuk melakukan beberapa segmentasi pelanggan dan optimalisasi strategi dalam penjualan (Kurniawan & Nugroho, 2025).

Selain itu, metode *K-Means Clustering* juga diterapkan untuk mengelompokkan produk ke dalam beberapa segmen berdasarkan kesamaan karakteristik. *K-Means* merupakan algoritma pemrograman yang berfungsi untuk mengelompokkan data menjadi beberapa grup (k) berdasarkan kesamaan ciri. Metode ini sering digunakan karena mudah diterapkan dan cukup cepat, terutama ketika digunakan untuk dataset yang sangat besar (Dhani et al., 2025). Melalui segmentasi produk, perusahaan dapat menemukan kelompok produk yang mempunyai penjualan tinggi, kelompok produk dengan harga tinggi, serta kelompok produk dengan penjualan rendah. Segmentasi ini bisa digunakan untuk merancang strategi pemasaran yang lebih terarah bagi setiap kelompok, seperti memberikan potongan harga untuk produk tertentu, meningkatkan promosi produk paling laris, atau menyesuaikan harga demi memperkuat daya saing. Selanjutnya *Elbow Method* adalah cara yang digunakan untuk mengidentifikasi jumlah *Cluster* terbaik pada analisis *K-Means*. Metode ini diterapkan untuk menentukan jumlah *Cluster* yang optimal dengan membandingkan jumlah *Cluster* yang membentuk siku pada grafik perbandingan SSE (Sum of Square Error) (Fitriyani & Jajuli, 2024). Proses pengolahan dan analisis data dilakukan dengan menggunakan *Google Colab*, yang memungkinkan penerapan algoritma secara interaktif serta mempermudah dokumentasi dan reproduksi hasil penelitian (Adam et al., 2025). Tujuan utama dari penelitian ini adalah untuk mengenali variabel-variabel yang memengaruhi penjualan produk Nike, menetapkan pola penjualan produk berdasarkan karakteristik spesifik, serta mengelompokkan produk ke dalam beberapa kluster dengan menggunakan metode *K-Means*. Studi ini juga bertujuan untuk mengidentifikasi produk yang mencapai volume penjualan tertinggi agar dapat dijadikan sebagai pedoman dalam strategi pemasaran dan pengembangan produk di waktu yang akan datang.

2. METODE PENELITIAN

2.1 Jenis dan Pendekatan Penelitian

Penelitian ini adalah penelitian kuantitatif yang menggunakan pendekatan eksploratif. Pendekatan ini diambil karena penelitian ini menekankan analisis data angka untuk mengidentifikasi pola penjualan, faktor penentu penjualan, dan segmentasi produk. Metode data mining dan pembelajaran mesin diterapkan untuk menciptakan analisis yang berdasarkan data dan objektif.

2.2 Sumber dan Jenis Data

Data yang dipakai diambil dari dataset penjualan produk Nike yang dapat diakses di platform Kaggle. Dataset ini memuat informasi mengenai produk, yang mencakup:

- Nama barang
- Biaya
- Penilaian pelanggan
- Produk kategori
- Total produk yang terjual

Buletin Ilmiah
Ilmu Komputer dan Multimedia
BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 3, No. 4, November 2025
ISSN 3024-8248 (media online)
Hal 550-557

Data bersifat sekunder dan digunakan dalam bentuk aslinya tanpa mengubah struktur utama. Dataset ini sangat cocok untuk riset yang memerlukan analisis penjualan dan klasifikasi berdasarkan fitur produk.

2.3 Tahapan Pengolahan Data

a. Pembersihan dan Persiapan Data

Tahap persiapan data adalah salah satu langkah terpenting untuk memastikan data berada dalam format yang optimal sebelum masuk ke tahap pemodelan (Belitung, 2025). Dalam penelitian ini melibatkan pemeriksaan nilai yang kosong, data yang tidak relevan, duplikasi data, serta kesalahan input dalam dataset. Proses ini bertujuan untuk memastikan bahwa semua data yang digunakan adalah valid dan siap untuk diproses oleh algoritma machine learning. Apabila terdapat ketidaksesuaian, perbaikan dilakukan seperti menghapus baris, mengisi nilai tertentu, atau melakukan normalisasi data.

b. Transformasi Data

Variabel angka seperti harga, penilaian, dan kuantitas terjual dijadikan sebagai input utama untuk model. Sementara itu, variabel kategori disesuaikan dengan keperluan analisis, khususnya dalam proses pengelompokan. Transformasi dilakukan agar setiap variabel berada dalam format yang sesuai dengan algoritma yang diterapkan.

c. Analisis Faktor Penentu Penjualan Menggunakan *Random Forest*

Random Forest diterapkan untuk menentukan variabel yang paling berpengaruh terhadap total penjualan produk Nike. Model dilatih dengan memanfaatkan fitur numerik, selanjutnya dilakukan analisis pentingnya fitur untuk menentukan urutan variabel berdasarkan tingkat dampaknya. Hasil ini dimanfaatkan untuk mengidentifikasi faktor-faktor utama yang memengaruhi pola penjualan.

d. Penentuan Jumlah *Cluster* Menggunakan *Elbow Method*

Elbow Method digunakan untuk menentukan jumlah *Cluster* yang paling tepat sebelum melaksanakan pengelompokan menggunakan *K-Means*. Langkah-langkah dilakukan dengan menghitung nilai SSE (Jumlah Kesalahan Kuadrat) dari berbagai nilai k, selanjutnya divisualisasikan dalam format grafik.

e. Pengelompokan Produk Menggunakan *K-Means*

Setelah ditemukan jumlah *Cluster* yang optimal, algoritma *K-Means* digunakan untuk mengelompokkan produk Nike berdasarkan kesamaan harga, penilaian, dan volume penjualan.

f. Visualisasi Data

Beberapa grafik dan plot dibuat di *Google Colab* untuk memperjelas hasil analisis, yang meliputi:

- Grafik *feature importance Random Forest*
- Grafik produk terlaris berdasarkan jumlah penjualan
- Grafik *Elbow Method*
- Scatter plot hasil *Clustering K-Means*

Buletin Ilmiah
Ilmu Komputer dan Multimedia
BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 3, No. 4, November 2025
ISSN 3024-8248 (media online)
Hal 550-557

Tabel 1. Dataset Sebelum Preprocessing

Invoice Date	*Product*	*Region*	*Retailer*	*Sales Method*	*State*	*Price Per Unit*	*Total Sales*	*Units Sold*
01-01-2020	*Men's Street Footwear*	*Northeast*	*Foot Locker*	*In-Store*	*New York*	50	6000	120
02-01-2020	*Men's Athletic Footwear*	*Northeast*	*Foot Locker*	*In-Store*	*New York*	50	5000	100
03-01-2020	*Women's Street Footwear*	*Northeast*	*Foot Locker*	*In-Store*	*New York*	40	4000	100
04-01-2020	*Women's Atgletic Footwear*	*Northeast*	*Foot Locker*	*In-Store*	*New York*	45	3825	85
05-01-2020	*Men's Apparel*	*Northeast*	*Foot Locker*	*In-Store*	*New York*	60	5400	90

Tabel 2. Dataset Setelah Preprocessing

Invoice Date	*Product*	*Region*	*Retailer*	*Sales Method*	*State*	*Price Per Unit*	*Total Sales*	*Units Sold*	*Retailer*	*Sales Method*	*State*
0	2	1	1	0	29	50	6000	120	1	0	29
24	1	1	1	0	29	50	5000	100	1	0	29
48	5	1	1	0	29	40	4000	100	1	0	29
72	4	1	1	0	29	45	3825	85	1	0	29
96	0	1	1	0	29	60	5400	90	1	0	29

3. ANALISA DAN PEMBAHASAN

Bagian ini menjelaskan hasil analisis dan visualisasi yang diperoleh dari penerapan algoritma *K-Means* dan *Random Forest* pada dataset penjualan produk Nike. Semua analisis dilaksanakan dengan menggunakan *Google Colab*.

3.1 Hasil Visualisasi Random Forest

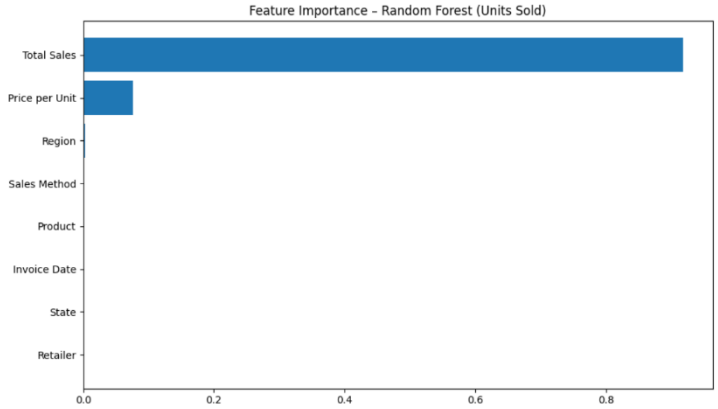
Analisis *Random Forest* dilakukan untuk menentukan variabel yang paling berpengaruh terhadap total penjualan. Visualisasi pentingnya fitur menunjukkan bahwa harga adalah variabel dengan kontribusi terbesar terhadap penjualan. Barang dengan harga lebih murah biasanya memiliki jumlah penjualan yang lebih tinggi. Hal ini menandakan bahwa konsumen sangat memperhatikan harga saat melakukan pembelian.

Variabel penilaian menduduki posisi kedua sebagai faktor penentu penjualan. Produk dengan peringkat tinggi mencerminkan kepuasan pelanggan yang baik, sehingga lebih banyak dicari. Sementara itu, variabel kategori memiliki dampak terendah pada jumlah penjualan karena variasi kategori tidak memberikan pengaruh signifikan terhadap minat beli. Melalui visualisasi *Random Forest*, dapat disimpulkan bahwa strategi harga dan kualitas produk berpengaruh besar terhadap kinerja penjualan Nike.

Visualisasi *feature importance* yang dihasilkan pada *Google Colab* memperlihatkan perbedaan yang signifikan antar variabel, sehingga dapat memudahkan dalam memahami hierarki faktor

Buletin Ilmiah
Ilmu Komputer dan Multimedia
BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 3, No. 4, November 2025
ISSN 3024-8248 (media online)
Hal 550-557

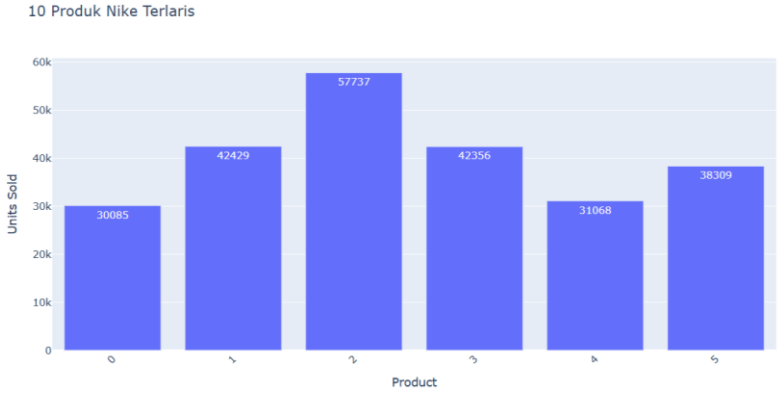
yang dapat memengaruhi penjualan produk. Hasil ini menunjukkan bahwa strategi harga dan peningkatan kualitas produk sangat penting dalam meningkatkan performa dalam penjualan produk Nike.



Gambar 1. Visualisasi feature importance Random Forest

3.2 Visualisasi Produk Terlaris

Google Colab menghasilkan grafik produk terlaris berdasarkan jumlah penjualan. Visualisasi ini memperlihatkan adanya beberapa produk yang memiliki jumlah penjualan jauh lebih tinggi dibandingkan produk lain. Produk-produk tersebut umumnya memiliki harga kompetitif dan rating yang baik. Hasil ini menunjukkan bahwa kombinasi harga terjangkau dan kualitas yang baik menjadi faktor utama produk masuk kategori *bestseller*. Informasi ini dapat menjadi acuan untuk menentukan strategi promosi, ketersediaan stok, dan pengembangan produk.

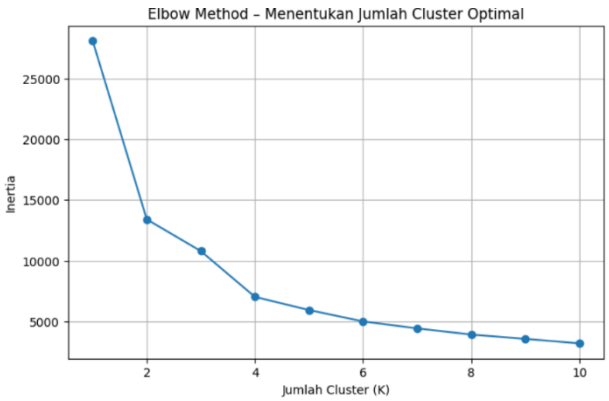


Gambar 2. Visualisasi 10 Produk Terlaris

3.3 Penentuan Jumlah Cluster Menggunakan Elbow Method

Elbow Method diterapkan untuk menentukan jumlah *Cluster* optimal sebelum proses *K-Means*. Grafik SSE yang dihasilkan menunjukkan titik siku pada k = 3**,** yang berarti tiga *Cluster* merupakan jumlah yang paling tepat untuk mewakili struktur data. Pemilihan jumlah *Cluster* ini digunakan sebagai dasar dalam proses pengelompokan, sehingga hasil *Clustering* dapat menggambarkan segmentasi produk secara maksimal.

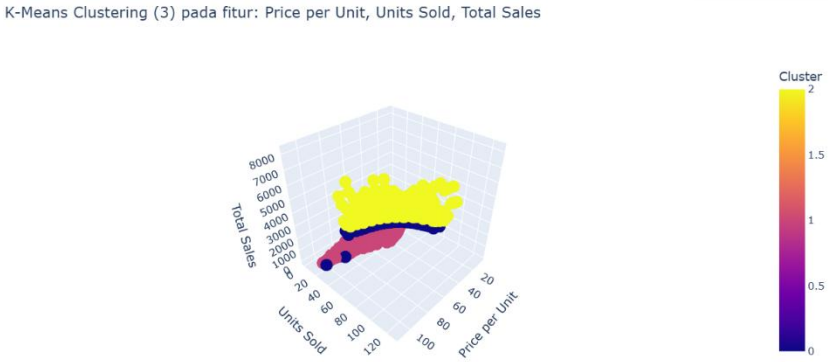
Buletin Ilmiah
Ilmu Komputer dan Multimedia
BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 3, No. 4, November 2025
ISSN 3024-8248 (media online)
Hal 550-557



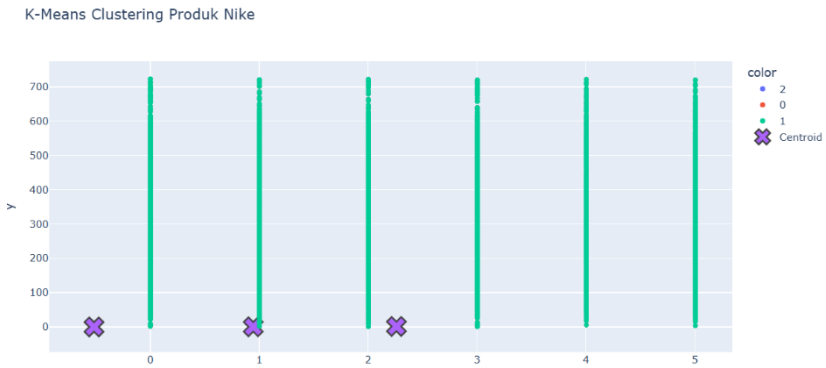
Gambar 3. Menetukan jumlah cluster menggunakan Elbow Method

3.4 Hasil *Clustering* Menggunakan *K-Means*

Cluster yang dihasilkan setelah proses *K-Means* adalah tiga kelompok . Masing-masing *Cluster* menunjukkan ciri-ciri yang berbeda sebagai berikut:



Gambar 4. Visualisasi *Clustering K-Means*



Gambar 5. Visualisasi Cluster K-Means dengan Centroid + Nama Produk

3.4.1 *Cluster* 1: Produk Harga Rendah dengan Penjualan Menengah

Produk di dalam *Cluster* ini memiliki harga yang lebih murah dibandingkan dengan *Cluster* lainnya. Penjualan berada pada posisi menengah, mengindikasikan bahwa harga yang rendah tidak selalu menghasilkan penjualan maksimal. Akan tetapi, *Cluster* ini tetap menunjukkan kinerja yang stabil dan menarik bagi pembeli yang peka terhadap harga.

Buletin Ilmiah
Ilmu Komputer dan Multimedia
BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 3, No. 4, November 2025
ISSN 3024-8248 (media online)
Hal 550-557

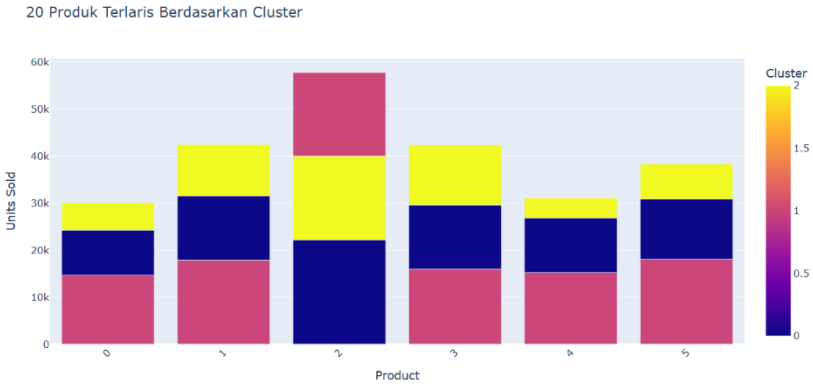
3.4.2 *Cluster* 2: Produk Premium dengan Penjualan Rendah

Cluster ini terdiri dari produk-produk yang memiliki harga tinggi dan volume penjualan yang rendah. Produk-produk di kategori ini ditujukan untuk pasar tertentu sehingga pangsa pasar yang dimiliki terbatas. Kenaikan harga membuat konsumen hanya membeli produk dalam keadaan tertentu.

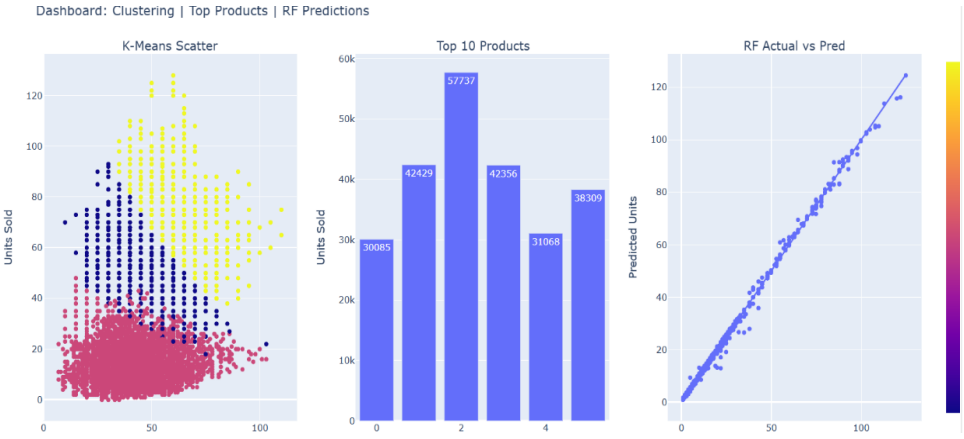
3.4.3 *Cluster* 3: Produk Harga Menengah dengan Penjualan Tinggi

Cluster ini memiliki ciri-ciri yang paling menarik karena mengandung produk dengan penjualan tertinggi. Produk dalam kelompok ini memiliki penilaian yang cukup baik dan harga yang sedang. Segmentasi ini mengindikasikan bahwa keseimbangan harga dan kualitas adalah kombinasi teroptimal untuk meningkatkan penjualan.

Visualisasi scatter plot menunjukkan pemisahan *Cluster* secara jelas dengan warna yang berbeda. Setiap kluster menunjukkan pola yang konsisten antara harga, penilaian, dan jumlah penjualan.



Gambar 6. Visualisasi 20 Produk terlaris berdasarkan Cluster



Gambar 7. Dashboard gabungan: *Clustering*, Top 10 *products, RF Actual vs Pred*

4. KESIMPULAN

Penelitian ini menghasilkan beberapa temuan yang signifikan terkait pola penjualan produk Nike melalui analisis *Random Forest* dan *K-Means*. Harga tersebut sudah terbukti sebagai faktor paling berpengaruh terhadap jumlah penjualan, sedangkan rating berperan sebagai faktor pendukung yang juga dapat memengaruhi minat beli. Hasil *Clustering* dapat mengindikasikan terdapat tiga kategori utama produk, yaitu produk dengan harga rendah dan penjualan menengah, produk premium yang memiliki penjualan rendah, serta produk harga menengah dengan penjualan tinggi.

Buletin Ilmiah
Ilmu Komputer dan Multimedia

BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 3, No. 4, November 2025
ISSN 3024-8248 (media online)
Hal 550-557

Segmentasi ini memberikan penjelasan yang tegas tentang ciri-ciri produk yang telah diinginkan oleh konsumen. Integrasi metode visualisasi dapat memperlihatkan hubungan antarvariabel dengan lebih mendalam dan dapat memberikan landasan solid bagi perusahaan dalam merumuskan strategi pemasaran, penetapan harga, serta pengembangan produk. Penelitian berikutnya bisa memasukkan variabel lain seperti *feedback* pelanggan atau informasi perilaku konsumen untuk mendapatkan hasil yang lebih menyeluruh.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada semua pihak yang telah berperan dalam proses penelitian ini, terutama kepada penyedia dataset, dosen pengampu, serta rekan-rekan yang memberikan bantuan dalam penyusunan naskah dan pelaksanaan analisis. Penulis juga menyampaikan penghargaan kepada platform *Google Colab* yang telah mendukung proses pengolahan data dan visualisasi dengan cara yang efisien.

REFERENCES

Adam, H., Novalia, E., & Hananto, A. L. (2025). *BULLETIN OF COMPUTER SCIENCE RESEARCH Prediksi Penjualan Barang Menggunakan Metode K-Means dan Regresi Linear*. *5*(4). https://doi.org/10.47065/bulletincsr.v5i4.541

Alfajr, N. H., & Defiyanti, S. (2024). *METODE RANDOM FOREST DAN PENERAPAN PRINCIPAL COMPONENT ANALYSIS (PCA)*. *12*(3).

Artikel, I., & Info, A. (2025). *ANALISIS SENTIMEN PENGGUNA APLIKASI SHOPEE PADA GOOGLE PLAY STORE MENGGUNAKAN METODE RANDOM FOREST*. *4*(3), 265–276.

Belitung, D. I. B. (2025). *ANALISIS DATA PELANGGAN DENGAN ALGORITMA K- MEANS UNTUK PENINGKATAN PENJUALAN LAYANAN ICONNET*. *2*(4), 50–60.

Dhani, A., Wardhani, R., Hanifah, A. I., Informatika, P. T., Lamongan, U. I., Lamongan, K., & Timur, J. (2025). *PENJUALAN MENGGUNAKAN K-MEANS SEBAGAI PENDUKUNG*. *6*(2).

Fitriyani, D., & Jajuli, M. (2024). *IMPLEMENTASI ALGORITMA K-MEANS UNTUK KLASTERISASI DALAM PENGELOLAAN PERSEDIAAN OBAT (STUDI KASUS : APOTEK NAZA)*. *12*(3), 2841–2848.

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

Kurniawan, S., & Nugroho, A. (2025). *E ISSN : 2809-4069 Analisis Faktor yang Mempengaruhi Promosi Karyawan Menggunakan Random Forest pada Dataset Employee Promotion*. *5*(2), 177–187.

Ramadhan, G., Faqih, A., Permana, F. E., Informatika, T., Informatika, M., & Cirebon, K. (2025). *ANALISIS TREN PENJUALAN MENU SEAFOOD DENGAN ALGORITMA*. *9*(4), 5942–5949.

Sari, A., Arifin, M., Darmanto, E., Teknik, F., & Kudus, U. M. (2025). *Prediksi kebutuhan stok barang menggunakan algoritma Random Forest untuk meningkatkan efisiensi penjualan*. *9*(2), 339–351.