

Model *Machine Learning* untuk Klasifikasi Warna *Fashion* Menggunakan Metode *K-Nearest Neighbor*

Yuke Manza¹, M. Suhada W.D.², Agus F.S. Ndruru³, Rika Rosnelly⁴
^{1,2,3,4}Program Studi Ilmu Komputer, Fakultas Teknik dan Ilmu Komputer,

Universitas Potensi Utama

¹manzayuke@gmail.com, ²muhammadsuhada191@gmail.com, ³agusndruru1999@gmail.com,

⁴rika@potensi-utama.ac.id

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan model klasifikasi warna *fashion* menggunakan metode *K-Nearest Neighbor* (KNN). Dengan mengumpulkan data dari 25.000 sampel yang mencakup variabel tinggi badan, berat badan, jenis kelamin, indeks massa tubuh, dan warna kulit, penelitian ini menerapkan tahapan metodologi yang sistematis, termasuk pengumpulan data, pra-pemrosesan, pemodelan, dan evaluasi. Hasil analisis menunjukkan bahwa model KNN mencapai akurasi sebesar 78.22%, dengan kemampuan yang baik dalam membedakan kelas positif dan negatif. Meskipun hasil ini memuaskan, terdapat peluang untuk meningkatkan akurasi model melalui optimasi parameter atau penggabungan dengan algoritma lain. Temuan ini memberikan kontribusi signifikan terhadap pengembangan teknologi *machine learning* dalam industri *fashion*, serta menawarkan solusi praktis bagi desainer dan produsen dalam memilih kombinasi warna yang menarik, sehingga dapat meningkatkan daya tarik produk di pasar.

Kata kunci: Machine learning, Klasifikasi, Warna, Fashion, K-nearest neighbor.

PENDAHULUAN

Fashion merupakan suatu ekosistem yang dinamis, terus berkembang, dan saling terhubung, yang menggabungkan unsur kreativitas dengan teknologi untuk menciptakan berbagai desain, gaya, dan banyak tren baru. *Fashion* selalu menjadi fitur penting dalam rutinitas kita sehari-hari. *Fashion* juga memainkan peran penting dalam kehidupan setiap orang. Pengetahuan tentang *fashion* mendorong orang untuk berpakaian dengan benar, tidak hanya memenuhi kebutuhan fisiologis penggunanya, tetapi juga kebutuhan praktik dan aktivitas sosial. Klasifikasi *fashion* tergantung pada karakteristiknya seperti warna, pola, tekstur, dan gaya. Rekomendasi *fashion* dibuat dengan menggunakan *feedback* dari pengguna, faktor harga, dan fitur gambar. Sehingga, pengenalan *fashion* adalah masalah yang beragam (Muhammed & Abdulazeez, 2024; Vijayaraj et al., 2022; Mameli et al., 2022; Jothi & Sulthana, 2021; Kolisnik et al., 2021).

Dalam penelitiannya, Muhammed dan Abdulazeez (2024) membahas tentang klasifikasi desain *fashion* berdasarkan *machine learning* dan *deep learning*, di mana dalam penelitian tersebut memberikan gambaran menyeluruh tentang penggunaan algoritma *machine learning* dalam klasifikasi desain *fashion*. Dalam penelitian lain, Li et al. (2020) menggunakan Jaringan Saraf Tiruan dengan metode *Convolutional Neural Networks* (CNN) untuk klasifikasi gambar *fashion* dan pencarian visual. Hasil menunjukkan akurasi tinggi pada klasifikasi sub-kategori produk hingga 96,6%. Sedangkan Vijayaraj et al. (2022) dalam penelitiannya melakukan klasifikasi desain *fashion* menggunakan *deep learning*, menunjukkan bahwa metode yang digunakan dapat

membedakan elemen pakaian dalam foto *fashion* dengan tingkat akurasi yang tinggi, sekitar 95,72% dalam pelatihan dan sekitar 94,52% dalam pengujian. Berdasarkan beberapa penelitian terdahulu, tantangan seperti keterbatasan data, dataset dengan resolusi rendah, ketidakseimbangan data, serta masalah *overfitting* pada model akan mempengaruhi tingkat akurasi dari model yang dibangun (Muhammed & Abdulazeez, 2024; Vijayaraj et al., 2022; Li et al., 2020).

Oleh karena itu, penulis mengusulkan model *machine learning* untuk klasifikasi warna *fashion*, yang dalam hal ini adalah pakaian (baju dan celana/rok) menggunakan metode *k-nearest neighbor* (KNN). *Machine learning* berkontribusi besar terhadap kemajuan *data analysis* dan *artificial intelligence*, *machine learning* merupakan algoritma yang dapat digunakan untuk membangun model yang dapat belajar tanpa diprogram secara eksplisit. KNN adalah salah satu metode *machine learning* yang paling banyak digunakan untuk klasifikasi. KNN termasuk metode klasifikasi klasik yang sederhana, kemampuan adaptasi dan kinerja yang baik, kemudahan pengaplikasian, populer, dan adaptif terhadap perubahan data. Metode KNN dapat diaplikasikan untuk klasifikasi sentimen, penyakit jantung, penyakit kulit, kanker paru, klasifikasi gambar, dan lain sebagainya (Wayahdi & Ruziq, 2022; Wayahdi et al., 2020; Araaf et al., 2023; Hutapea & Silalahi, 2023; Wibisono et al., 2023).

Tujuan penelitian ini adalah mengembangkan model *machine learning* menggunakan metode *k-nearest neighbor* untuk mengklasifikasikan warna *fashion* pakaian sehingga memberikan rekomendasi warna yang dapat membantu desainer dan produsen dalam memilih warna yang tepat untuk produk mereka, serta membantu masyarakat umum untuk menentukan kombinasi warna pakaian yang sesuai untuk digunakan. Penulis juga akan melakukan analisis pada akurasi dan melakukan evaluasi pada model yang dibangun. Dalam industri *fashion*, pemilihan warna merupakan salah satu faktor penting yang mempengaruhi daya tarik produk. Harapannya, metode *k-nearest neighbor* dapat digunakan untuk menganalisis data warna dan memberikan rekomendasi warna *fashion* yang sesuai.

METODE PENELITIAN

Penelitian ini dilakukan dengan beberapa tahapan metodologi yang sistematis untuk memastikan bahwa model yang dibangun dapat memberikan hasil prediksi yang akurat dan dapat diandalkan. Tahapan tersebut meliputi *data collection*, *data preprocessing*, *modeling*, dan *evaluation*. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

Gambar 1 menunjukkan tahapan penelitian yang dilakukan, yaitu:

1. *Data Collection* (Pengumpulan Data)

Tahapan pertama dalam penelitian ini adalah pengumpulan data. Dataset yang digunakan dalam penelitian ini diperoleh dari dataset yang telah dipublikasikan. Dataset ini mencakup berbagai fitur yang relevan, yaitu tinggi badan, berat badan, jenis kelamin, indeks masa tubuh, dan warna kulit dengan total 5 variabel, serta 1 label (target) yaitu warna *fashion*. Data yang dikumpulkan mencakup 25.000 data yang mencerminkan berbagai karakteristik orang. Pengumpulan data yang komprehensif ini bertujuan untuk memastikan bahwa model dapat belajar dari berbagai skenario dan kondisi yang mungkin ditemui dalam proses pemilihan atau klasifikasi warna *fashion* pakaian. Gambar 2 menunjukkan dataset yang digunakan, menampilkan lima data teratas.

```
df = pd.read_excel('/content/drive/MyDrive/Dataset/Profile-of-Body-Metrics-and-Fashion-Colors.xlsx')
df.head()
```

	Height(Centimeter)	Weight(Kilograms)	Gender	BMI	Skin Color	Fashion Color
0	167.089607	51.252494	Female	Underweight	176.000000	GG
1	181.648633	61.909547	Male	Ideal	152.333333	BB
2	176.272800	69.411778	Male	Ideal	176.000000	BB
3	173.270164	64.562199	Male	Ideal	85.000000	GG
4	172.181037	65.452010	Male	Ideal	176.000000	GR

Gambar 2. Dataset yang Digunakan

2. Data Preprocessing (Pra-pemrosesan Data)

Pada tahap pra-pemrosesan, data yang diperoleh dibersihkan dan dipersiapkan (*cleaned and prepared*) untuk persiapan proses selanjutnya. Proses ini mencakup beberapa langkah, seperti mengatasi nilai yang hilang, menghapus duplikasi, dan mengubah variabel kategorikal ke dalam format numerik. Selain itu, fitur-fitur yang tidak relevan atau redundan juga dihilangkan untuk meningkatkan efisiensi model. Normalisasi data juga dilakukan untuk memastikan bahwa semua fitur berada dalam skala yang sama, sehingga tidak ada fitur yang mendominasi proses pelatihan model. Selanjutnya, dataset dibagi menjadi dua bagian, yaitu data latih dan data uji dengan proporsi 80% dan 20%.

3. Modelling (Pemodelan)

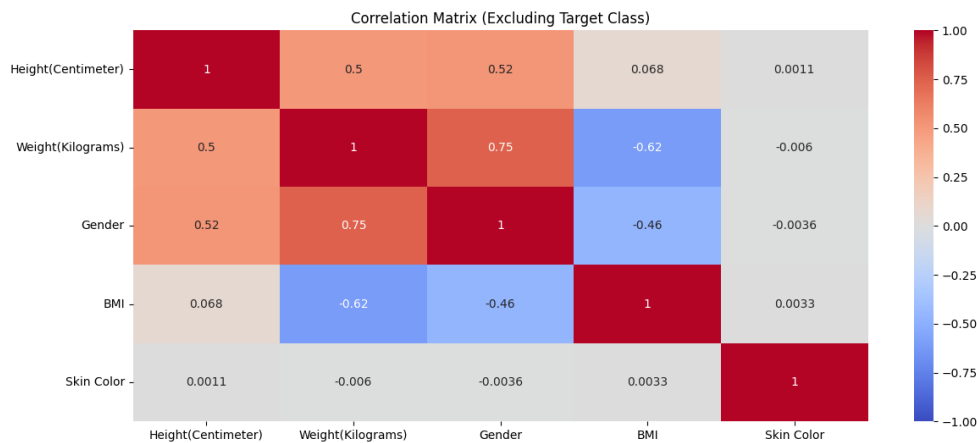
Pada tahap pemodelan, algoritma *K-Nearest Neighbor* (KNN) dipilih untuk membangun model klasifikasi warna fashion. Model KNN digunakan karena cukup simpel serta memiliki kemampuan yang baik dalam menangani data yang tidak terstruktur. Proses pelatihan dilakukan dengan menggunakan data *training*, di mana model dilatih untuk mengenali pola dan hubungan antara fitur dan target warna fashion.

4. Evaluation (Evaluasi)

Model yang telah dibangun, dievaluasi dengan menggunakan data uji untuk mengukur kinerjanya. Beberapa metrik evaluasi yang digunakan dalam penelitian ini antara lain *accuracy*, *precision*, *recall*, dan *F1-score*. Metrik-metrik ini memberikan gambaran secara menyeluruh mengenai seberapa baik model dalam melakukan klasifikasi warna fashion. Selain itu, tahap evaluasi juga dilakukan untuk menilai kemampuan model dalam membedakan kelas positif dan negatif.

HASIL DAN PEMBAHASAN

Hasil penelitian ditampilkan dalam bentuk matriks korelasi, yang menggambarkan hubungan antara berbagai fitur dalam dataset yang digunakan untuk analisis klasifikasi warna *fashion*. Matriks ini memberikan wawasan penting mengenai interaksi antar fitur yang dapat mempengaruhi keputusan dalam proses klasifikasi. Dengan menganalisa nilai korelasi, kita dapat mengidentifikasi pola dan hubungan yang signifikan, dan memahami bagaimana setiap fitur berkontribusi terhadap hasil klasifikasi. Nilai korelasi berkisar antara -1 hingga 1, di mana nilai positif menunjukkan hubungan langsung dan nilai negatif menunjukkan hubungan terbalik. Gambar 3 menampilkan matriks korelasi yang dihasilkan dari penelitian ini.

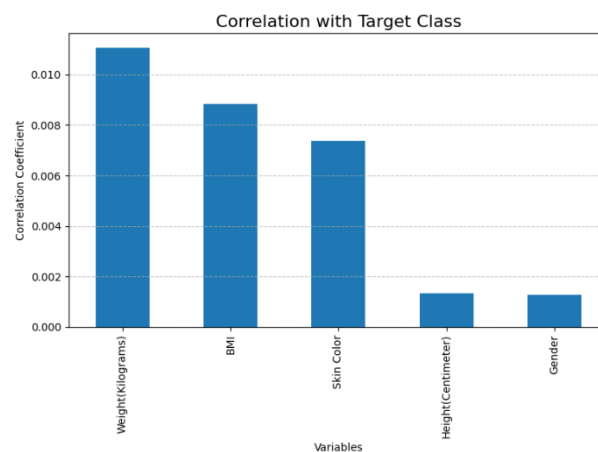


Gambar 3. Matriks Korelasi Antar Variabel

Pada Gambar 3 tentang matriks korelasi antar variabel dapat disimpulkan bahwa:

1. Tinggi dan Berat: Terdapat korelasi positif sebesar 0.5, menunjukkan hubungan yang cukup kuat antara tinggi badan dan berat badan.
2. Tinggi dan Jenis Kelamin: Terdapat korelasi positif sebesar 0.52, yang juga menunjukkan hubungan yang cukup kuat antara tinggi badan dan jenis kelamin.
3. Berat dan Jenis Kelamin: Terdapat korelasi positif tertinggi yaitu sebesar 0.75, yang menunjukkan hubungan yang kuat antara berat badan dan jenis kelamin.
4. Jenis Kelamin dan BMI: Korelasi negatif sebesar -0.46, yang menunjukkan bahwa ada hubungan terbalik antara jenis kelamin dan indeks masa tubuh.
5. BMI dan Berat: Korelasi negatif sebesar -0.62, yang menunjukkan bahwa ada hubungan terbalik antara indeks masa tubuh dan berat.
6. Warna Kulit: Korelasi dengan variabel lain sangat rendah, menunjukkan bahwa warna kulit tidak memiliki hubungan yang signifikan dengan variabel lainnya dalam konteks ini.

Secara keseluruhan, matriks ini memberikan gambaran tentang bagaimana variabel-variabel tersebut saling berhubungan satu sama lain. Matriks korelasi ini juga memberikan wawasan yang berharga untuk memahami interaksi antara fitur-fitur dalam dataset, yang dapat membantu dalam pengembangan model klasifikasi warna *fashion* yang lebih akurat. Namun, korelasi antara variabel dan target juga perlu diperhatikan, Gambar 4 menunjukkan korelasi variabel dengan kelas target.

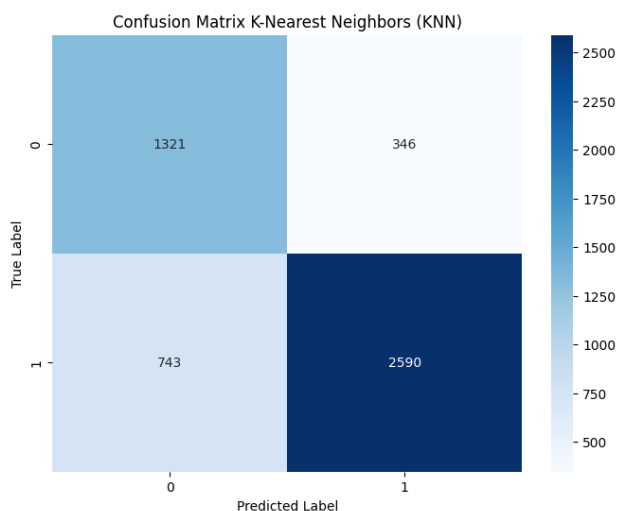


Gambar 4. Korelasi Variabel dengan Kelas Target

Gambar 4 menampilkan diagram yang menggambarkan koefisien korelasi antara beberapa variabel dan kelas target. Koefisien korelasi ini menunjukkan seberapa kuat hubungan antara masing-masing variabel dengan kelas target. Pada Gambar 4 dapat disimpulkan bahwa:

1. Berat: Memiliki koefisien korelasi tertinggi sekitar 0.010, menunjukkan bahwa berat badan memiliki hubungan yang paling signifikan dengan kelas target dibandingkan variabel lainnya.
2. BMI (indeks massa tubuh): Menunjukkan korelasi yang cukup signifikan dengan nilai sekitar 0.008, menandakan bahwa BMI juga memiliki pengaruh terhadap kelas target.
3. Warna Kulit: Memiliki koefisien korelasi sekitar 0.006, menunjukkan adanya hubungan meskipun tidak sekuat berat badan dan BMI.
4. Tinggi dan Jenis Kelamin: Kedua variabel ini memiliki koefisien korelasi yang sangat rendah, menunjukkan bahwa variabel ini memiliki pengaruh yang minimal terhadap kelas target.

Hasil pemodelan *machine learning* untuk klasifikasi warna *fashion* dengan algoritma *K-Nearest Neighbor* menunjukkan hasil yang cukup baik. Gambar 5 menunjukkan hasil analisis algoritma *K-Nearest Neighbor* dalam bentuk *confusion matrix*.



Gambar 5. Analisis *Confusion Matrix*

Hasil kinerja dari model klasifikasi *K-Nearest Neighbors* (KNN) berdasarkan metrik evaluasi yang umum digunakan dalam analisis klasifikasi dapat dilihat pada Tabel 1.

Tabel 1. Hasil Evaluasi Model KNN

Model	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
K-Nearest Neighbor	2590	1321	346	743	78.22%	88.2%	77.7%	82.4%

Dari analisis *confusion matrix* (Gambar 5) dan evaluasi model (Tabel 1), dapat disimpulkan bahwa model *K-Nearest Neighbor* menunjukkan performa yang cukup baik dalam hal klasifikasi warna *fashion*, di mana model *K-Nearest Neighbor* memperoleh akurasi 78.22%. Selain itu, model ini juga mencatat nilai akurasi, presisi, recall, dan F1-score yang cukup baik. Hal ini menunjukkan bahwa model *K-Nearest Neighbor* cukup efektif dalam melakukan klasifikasi kelas positif dan negatif, yang dapat berimplikasi positif pada proses pengambilan keputusan.

KESIMPULAN

Penelitian ini berhasil mengembangkan model klasifikasi warna *fashion* menggunakan metode *K-Nearest Neighbor* (KNN), yang menunjukkan akurasi sebesar 78.22%. Melalui tahapan metodologi yang sistematis, penelitian ini mengidentifikasi pola signifikan antara variabel-variabel seperti berat badan dan kelas target, memberikan wawasan penting untuk pengembangan model yang lebih akurat dalam industri *fashion*. Hasil evaluasi model menunjukkan bahwa KNN efektif dalam membedakan kelas positif dan negatif, serta dapat diandalkan dalam memberikan rekomendasi warna yang tepat. Meskipun demikian, terdapat potensi untuk mengoptimasi model KNN agar mencapai akurasi yang lebih tinggi, baik melalui penyempurnaan parameter model maupun dengan menggabungkan KNN dengan algoritma lain. Temuan ini tidak hanya berkontribusi pada kemajuan teknologi *machine learning* dalam klasifikasi *fashion*, tetapi juga menawarkan solusi praktis bagi desainer dan produsen dalam memilih kombinasi warna yang menarik, sehingga meningkatkan daya tarik produk di pasar.

DAFTAR PUSTAKA

- Araaf, M. A., Nugroho, K., & Setiadi, D. R. I. M. (2023). Comprehensive Analysis and Classification of Skin Diseases based on Image Texture Features using K-Nearest Neighbors Algorithm. *Journal of Computing Theories and Applications*, vol. 1, no. 1, pp. 31–40. doi: 10.33633/jcta.v1i1.9185
- Hutapea, M. I. & Silalahi, A. P. (2023). Moderna's Vaccine Using the K-Nearest Neighbor (KNN) Method: An Analysis of Community Sentiment on Twitter. *Jurnal Penelitian Pendidikan IPA*, vol. 9, no. 5, pp. 3808–3814. doi: 10.29303/jppipa.v9i5.3203
- Jothi, J. A. A. & Sulthana, A. R. (2021). A review on the literature of fashion recommender system using deep learning. *Totem Publishers Ltd*. doi: 10.23940/ijpe.21.08.p5.695702
- Kolisnik, B., Hogan, I., & Zulkernine, F. (2021). Condition-CNN: A hierarchical multi-label fashion image classification model. *Expert Syst Appl*, vol. 182. doi: 10.1016/j.eswa.2021.115195.
- Li, F., Kant, S., Araki, S., Bangera, S. & Shukla, S. S. (2020). Neural Networks for Fashion Image Classification and Visual Search. [Online]. Available: <http://arxiv.org/abs/2005.08170>
- Mameli, M., Paolanti, M., Pietrini, R., Pazzaglia, G., Frontoni, E., & Zingaretti, P. (2022). Deep Learning Approaches for Fashion Knowledge Extraction from Social Media: A Review. *Institute of Electrical and Electronics Engineers Inc*. doi: 10.1109/ACCESS.2021.3137893
- Muhammed, A. A. & Abdulazeez, A. M. (2024). Fashion Design Classification Based on Machine Learning and Deep Learning Algorithms: A Review. *Indonesian Journal of Computer Science*
- Vijayaraj, A. et al. (2022). Deep Learning Image Classification for Fashion Design. *Wireless Communications and Mobile Computing*. doi: 10.1155/2022/7549397
- Wayahdi, M. R. & Ruziq, F. (2022). KNN and XGBoost Algorithms for Lung Cancer Prediction. *Journal of Science Technology (JoSTec)*, vol. 4, no. 1, pp. 179–186. doi: 10.55299/jostec.v4i1.251
- Wayahdi, M. R., Syahputra, D., & Ginting, S. H. N. (2020). Evaluation of the K-Nearest Neighbor Model with K-Fold Cross Validation on Image Classification. *Infokum*
- Wibisono, A. D. R., Hidayat, S., Ramadhan, H. M. T., & Puspaningrum, E. Y. (2023). Comparison of K-Nearest Neighbor and Decision Tree Methods using Principal Component Analysis Technique in Heart Disease Classification. *Indonesian Journal of Data and Science*, vol. 4, no. 2, pp. 90–100. doi: 10.56705/ijodas.v4i2.70