



IDENTIFIKASI KATEGORI PEROKOK PADA CITRA WAJAH SANTRI MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR* DAN LINEAR DISCRIMINANT ANALYSIS

Ali Nurdiansyah¹, Sumijan², Gunadi Widi Nurcahyo.³

^{1,2,3}Magister Teknik Informatika, FILKOM UPI “YPTK” Padang

^{1,2,3}Jl. Raya Lubuk Begalung, Kota Padang, 25148

E-mail : alinurdiansyah30@gmail.com¹, sumijan@upiyptk.ac.id², gwidinurcahyo@gmail.com³

ABSTRAK

Merokok telah menjadi salah satu kebiasaan yang lazim dilakukan dalam kehidupan sehari-hari di berbagai negara, termasuk Indonesia. Laporan World Health Organization (WHO), menjelaskan bahwa tembakau telah menyebabkan lebih dari 5 juta kematian pada tahun 2020, dengan sekitar 70% korban berasal dari negara berkembang, terutama di wilayah Asia khususnya Indonesia. Berdasarkan hal tersebut maka, penelitian ini bertujuan untuk mengidentifikasi kategori perokok pada citra wajah santri menggunakan algoritma *K-Nearest Neighbor* (*K-NN*) dan *Linear Discriminant Analysis* (*LDA*). Metode penelitian menggunakan *K-Nearest Neighbor* (*K-NN*) dan *Linear Discriminant Analysis* (*LDA*). Dataset penelitian ini menggunakan data dari wajah santri dan situs www.kaggle.com yang digunakan sebanyak 70 citra yang terdiri dari 50 data primer dan 20 data sekunder. Hasil dari pengujian didapatkan bahwa citra wajah santri yang di ekstrak menggunakan algoritma *K-NN* dan *LDA* mendapatkan hasil akurasi sebanyak 92,5% yang terdiri dari 70% data latih dan 30% data uji. Kontribusi dalam penelitian ini memberikan manfaat terhadap pengembangan aplikasi machine learning untuk deteksi perilaku menyimpang berbasis data dan membantu pihak pesantren dalam mendeteksi serta menangani santri yang terindikasi merokok dengan pendekatan teknologi.

Kata kunci : Merokok, Machine Learning, Citra Digital, *K-NN*, *LDA*

ABSTRACT

Smoking has become a common habit in daily life in various countries, including Indonesia. A report by the WorldHealth Organization (WHO) report explains that tobacco caused more than 5 million deaths in 2020, with around 70% of victims coming from developing countries, especially in Asia, particularly Indonesia. Based on this, this study aims to identify categories of smokers in images of santri faces using the *K-Nearest Neighbor* (*K-NN*) and *Linear Discriminant Analysis* (*LDA*) algorithms. The research method used *K-Nearest Neighbor* (*K-NN*) and *Linear Discriminant Analysis* (*LDA*). The research dataset used data from santri faces and the website www.kaggle.com, consisting of 70 images comprising 50 primary data and 20 secondary data. The results of the testing showed that the images of students' faces extracted using the *K-NN* and *LDA* algorithms achieved an accuracy of 92.5%, consisting of 70% training data and 30% test data. The contribution of this research provides benefits for the development of machine learning applications for detecting deviant behavior based on data and assists Islamic boarding schools in detecting and handling students who are suspected of smoking using a technological approach.

Keywords: Smoking, Machine Learning, Digital Imaging, *K-NN*, *LDA*

1. PENDAHULUAN

Merokok telah menjadi salah satu kebiasaan

yang lazim dilakukan dalam kehidupan sehari-hari di berbagai negara, termasuk Indonesia (Vladea Rahman 2023). Faktanya sebagian besar



masyarakat sudah mengetahui bahaya merokok, perilaku merokok masih menjadi kebudayaan yang sulit diubah (Nurlina 2024). Menurut laporan dari World Health Organization (WHO), tembakau telah menyebabkan lebih dari 5 juta kematian pada tahun 2020, dengan sekitar 70% korban berasal dari negara berkembang, terutama di wilayah Asia (Gunarsih, Wiyono, and Af'idah 2024). Indonesia sebagai salah satu negara dengan tingkat konsumsi rokok tertinggi di dunia, menempati peringkat ke-5 setelah China, Amerika Serikat, Jepang, dan Rusia (Gunarsih, Wiyono, and Af'idah 2024).

Perilaku merokok pada remaja, termasuk santri baru di lingkungan pesantren, menjadi fenomena sosial yang memprihatinkan karena bertentangan dengan nilai-nilai religius dan kedisiplinan yang diajarkan di lembaga pendidikan tersebut (Khirunisah 2024). Pesantren seharusnya menjadi lingkungan yang steril dari kebiasaan buruk seperti merokok, namun dalam kenyataannya, praktik merokok secara diam-diam masih terjadi di kalangan santri (Alim 2024). Faktor-faktor yang memengaruhi perilaku ini tidak hanya berasal dari pengaruh internal seperti rasa ingin tahu atau pencarian jati diri, tetapi juga dipicu oleh pengaruh eksternal seperti tekanan teman sebaya dan lemahnya pengawasan lingkungan (Lubis et al. 2025).

Penelitian Riyadi dan Wahyu (2024) mengungkapkan bahwa sebanyak 89% santri di salah satu pesantren menunjukkan kecenderungan merokok pada kategori sedang, dan 78% dari mereka mengalami pengaruh teman sebaya yang signifikan, yang menjadi salah satu faktor pendorong utama perilaku tersebut (Isnaini 2023). Kemajuan teknologi dalam bidang kecerdasan buatan dan *machine learning* memberikan solusi baru dalam proses deteksi perilaku (Rifai, Pramudya, and Narfandi 2024). Studi yang dilakukan oleh Rokhmah dkk menjelaskan bahwa perilaku merokok pada santri tidak dapat dilepaskan dari tiga komponen utama, yakni faktor predisposisi (pengetahuan, sikap, dan motivasi pribadi), faktor pendukung (aksesibilitas terhadap rokok dan lemahnya pengawasan), serta faktor penguat (dukungan sosial dari teman sebaya atau figur yang merokok) (Haikhah 2021). Aspek psikologis juga memainkan peran penting dalam membentuk perilaku merokok. Penelitian oleh Rahmadanty dkk menunjukkan bahwa sebanyak 47 santri laki-laki terindikasi menggunakan zat adiktif termasuk rokok, dan perilaku ini berkorelasi dengan masalah psikososial seperti stres, kecemasan, dan kebutuhan akan pelarian emosional (Dwinda 2025). Persepsi santri tentang

merokok masih diliputi oleh pemahaman keliru, seperti anggapan bahwa merokok adalah simbol kedewasaan atau wujud eksistensi diri, yang semakin diperkuat oleh lemahnya regulasi dan pengaruh media (Dwinda 2025).

Penelitian Riyadi dan Wahyu (2024) mengungkapkan bahwa sebanyak 89% santri di salah satu pesantren menunjukkan kecenderungan merokok pada kategori sedang, dan 78% dari mereka mengalami pengaruh teman sebaya yang signifikan, yang menjadi salah satu faktor pendorong utama perilaku tersebut (Isnaini 2023). Kemajuan teknologi dalam bidang kecerdasan buatan dan *machine learning* memberikan solusi baru dalam proses deteksi perilaku (Rifai, Pramudya, and Narfandi 2024). Studi yang dilakukan oleh Rokhmah dkk menjelaskan bahwa perilaku merokok pada santri tidak dapat dilepaskan dari tiga komponen utama, yakni faktor predisposisi (pengetahuan, sikap, dan motivasi pribadi), faktor pendukung (aksesibilitas terhadap rokok dan lemahnya pengawasan), serta faktor penguat (dukungan sosial dari teman sebaya atau figur yang merokok) (Haikhah 2021). Aspek psikologis juga memainkan peran penting dalam membentuk perilaku merokok. Penelitian oleh Rahmadanty dkk menunjukkan bahwa sebanyak 47 santri laki-laki terindikasi menggunakan zat adiktif termasuk rokok, dan perilaku ini berkorelasi dengan masalah psikososial seperti stres, kecemasan, dan kebutuhan akan pelarian emosional (Dwinda 2025). Persepsi santri tentang merokok masih diliputi oleh pemahaman keliru, seperti anggapan bahwa merokok adalah simbol kedewasaan atau wujud eksistensi diri, yang semakin diperkuat oleh lemahnya regulasi dan pengaruh media (Dwinda 2025).

Implementasi dari konsep ini sesuai dengan penerapan *machine learning* (ML) merupakan cabang dari kecerdasan buatan yang memungkinkan komputer untuk belajar dari data dan meningkatkan kinerjanya tanpa perlu diprogram secara eksplisit (Maulani et al. 2025). Proses pembelajaran ini dilakukan dengan mengekstraksi pola atau pengetahuan dari data sehingga sistem mampu melakukan prediksi atau pengambilan keputusan secara lebih akurat (Febrianto, Wijaya, and Rilvani 2025). Metode ML secara umum dapat dibedakan menjadi tiga kategori utama, yaitu supervised learning yang memanfaatkan data berlabel untuk pelatihan, unsupervised learning yang mencari pola pada data tanpa label, dan reinforcement learning yang memanfaatkan interaksi dengan lingkungan serta umpan balik berupa reward atau punishment untuk



memperbaiki performa (Muflikhah, Mahmudy, and Kurnianingtyas 2023).

Metode yang digunakan dalam konsep machine learning dapat mengadopsi kinerja metode *K-Nearest Neighbor* (K-NN). *K-Nearest Neighbor* (KNN), yang dikenal efektif dalam mengenali pola dari data perilaku (Akbar 2024). Penelitian Identifikasi Perilaku Merokok pada Santri Baru, K-NN dimanfaatkan untuk mengklasifikasikan citra wajah santri ke dalam kategori “perokok” atau “bukan perokok” berdasarkan ciri-ciri visual yang telah diekstraksi.

Metode yang terlibat dalam *machine learning* juga dapat dilihat dari performa metode Linear Discriminant Analysis (LDA). LDA dalam proses klusterisasi, seperti yang dilakukan oleh Surianto dan Surianto (2025), mampu meningkatkan kualitas pembentukan kelompok data secara signifikan (Karim 2025). LDA meningkatkan metrik evaluasi klasifikasi (*precision*, *recall*, *F1-score*) dibandingkan data asli tanpa reduksi (Rahmawati, Wahyuni, and Kartika 2025).

Riwayat penelitian terdahulu terhadap metode penelitian dalam menyelesaikan masalah klasifikasi citra sayuran, khususnya kol hijau, kol ungu, dan wortel, menggunakan kombinasi metode Linear Discriminant Analysis (LDA) untuk ekstraksi ciri dan *K-Nearest Neighbor* (K-NN) untuk klasifikasi (Nurdiansyah et al. 2025). Objek penelitian berupa 58 citra sayuran yang diambil menggunakan kamera beresolusi tinggi, kemudian dibagi menjadi 30 citra untuk data latih dan 28 citra untuk data uji. Metode K-NN digunakan untuk mengklasifikasikan citra berdasarkan jarak terdekat antara data uji dan data latih. Hasil pengujian menunjukkan tingkat akurasi klasifikasi sebesar 92,8571%, dengan waktu pemrosesan rata-rata 8–10 detik per citra dan konsistensi kinerja tinggi meskipun kondisi pencahayaan bervariasi (Meldyantono 2025).

Penelitian yang dilakukan oleh Kosasih bertujuan untuk mengembangkan sistem pengenalan wajah menggunakan metode Linear Discriminant Analysis (LDA) sebagai teknik ekstraksi fitur dan *K-Nearest Neighbor* (K-NN) sebagai metode klasifikasi, dengan memperhatikan pengaruh jumlah data latih terhadap akurasi pengenalan. Objek penelitian berupa 400 citra wajah dari 40 orang, masing-masing memiliki 10 citra dengan variasi ekspresi yang diambil dari basis data ORL. Data dibagi ke dalam beberapa skenario, yaitu jumlah data latih per orang sebanyak 5, 6, 7, 8, dan 9 citra, sementara sisanya digunakan sebagai data uji. Hasil pengujian menunjukkan bahwa akurasi tertinggi sebesar

97,5% dicapai ketika jumlah data latih per orang adalah 9 citra (total 360 data latih) dan jumlah tetangga terdekat (K) adalah 1. Secara umum, akurasi menurun ketika jumlah data latih berkurang atau nilai K semakin besar, menegaskan bahwa proporsi data latih yang lebih banyak dan pemilihan K yang tepat berpengaruh signifikan terhadap performa sistem (Khaliqah, Sarifah, and Khotijah 2024).

Penelitian yang dilakukan oleh Yulianti dan Arifin (2017) bertujuan untuk membangun sistem identifikasi jenis ikan hias air tawar menggunakan metode *K-Nearest Neighbor* (K-NN). Objek penelitian berupa citra beberapa jenis ikan hias air tawar, seperti ikan mas koki, ikan louhan, ikan koi, ikan komet, dan ikan arwana, yang diambil menggunakan kamera digital. Hasil pengujian menunjukkan bahwa nilai K yang tepat dapat meningkatkan akurasi klasifikasi, di mana akurasi tertinggi mencapai 86,67% pada kondisi tertentu. Penelitian ini membuktikan bahwa metode K-NN efektif untuk mengklasifikasikan objek visual yang memiliki variasi bentuk dan warna, serta dapat diadaptasi untuk aplikasi identifikasi objek lain berbasis citra digital (Yusmah Sari and Purwanti Ningrum 2017).

Berdasarkan penjelasan sebelumnya, penelitian ini akan melakukan proses identifikasi kategori perokok pada citra wajah santri. Proses identifikasi dikembangkan dengan memanfaatkan kinerja algoritma *K-Nearest Neighbor* (K-NN) dan Linear Discriminant Analysis (LDA) untuk menyajikan klasifikasi secara otomatis. Pengembangan metode klasifikasi disajikan dalam bentuk kombinasi algoritma K-NN dan LDA guna meningkatkan akurasi identifikasi. Algoritma K-NN bekerja dengan memanfaatkan kemiripan fitur wajah, sementara LDA berperan dalam reduksi dimensi dan pemisahan kelas secara optimal. Performa kombinasi kedua algoritma ini diharapkan dapat memberikan hasil identifikasi yang tepat dan akurat. Penelitian ini juga akan menyajikan model identifikasi yang efektif dan efisien dalam menentukan kategori perokok berdasarkan citra wajah santri. Kontribusi penelitian ini diharapkan dapat berdampak pada pengembangan sistem identifikasi berbasis citra, khususnya dalam bidang kesehatan dan pemantauan gaya hidup di lingkungan pesantren. Hasil penelitian juga dapat menjadi referensi bagi pengembangan teknik pengolahan citra dan *machine learning* untuk aplikasi serupa.



2. METODE PENELITIAN

Metode penelitian merupakan tahapan-tahapan sistematis yang dilakukan oleh penulis dalam menyelesaikan penelitian yang berhubungan dengan menentukan prioritas manajemen risiko menggunakan metode *K-Nearest Neighbor* (K-NN) dan *Linear Discriminant Analysis* (LDA) secara umum yang dapat digambarkan melalui *flowchart* yang dapat dilihat pada Gambar 1. yang menunjukkan langkah-langkah yang ditempuh dari awal hingga akhir untuk mencapai hasil yang optimal.

A. Pengumpulan Citra

Peneliti melakukan proses pengumpulan data. Dataset yang digunakan dalam studi ini terdiri atas gambar wajah santri laki-laki kelas tujuh dan sepuluh Pondok Pesantren Kauman Muhammadiyah Padang Panjang yang diambil menggunakan kamera digital dalam format .JPG. Peneliti kemudian melakukan tahap pembuatan dataset dengan melakukan pre-processing, yaitu mengubah ukuran citra menjadi 100x100 piksel, mengganti latar belakang, dan mengubah jenis file citra menjadi JPG. Dari 50 data citra yang tersedia, hasil pre-processing menghasilkan dua jenis data, yaitu data latih dan data uji. Terdapat sebanyak 30 data citra untuk data latih dan 20 data citra untuk data uji, yang masing-masing dikategorikan ke dalam tiga kelas, yaitu tidak perokok dan perokok aktif.

B. Pre-processing

Proses *pre-processing* citra merupakan tahapan awal yang sangat penting dalam sistem identifikasi perilaku merokok berbasis citra wajah. Tahapan ini bertujuan untuk mempersiapkan citra agar siap digunakan pada proses ekstraksi fitur dan klasifikasi. Langkah pertama dalam pre-processing adalah input citra, yaitu memasukkan gambar wajah santri baru ke dalam sistem, baik melalui pengambilan langsung menggunakan kamera maupun dari dataset yang telah tersedia. Selanjutnya, dilakukan proses *crop dan resizing*, yaitu pemotongan area wajah agar fokus hanya pada bagian yang relevan, serta penyesuaian ukuran citra menjadi seragam untuk memudahkan proses pengolahan selanjutnya.

Setelah itu, citra yang telah dicrop dan diubah ukurannya menjadi 100x100 piksel akan dinormalisasi. Normalisasi ini dilakukan untuk menyeragamkan nilai intensitas piksel, biasanya dalam rentang 0 sampai 1, agar sistem tidak terpengaruh oleh perbedaan pencahayaan pada tiap citra. Langkah berikutnya adalah konversi warna ke *grayscale*, yaitu mengubah citra berwarna menjadi

hitam-putih untuk menyederhanakan data, mengingat informasi warna tidak dibutuhkan dalam proses identifikasi perilaku merokok. Kemudian dilakukan *noise reduction*, yaitu penghilangan gangguan atau bintik acak pada citra yang dapat memengaruhi akurasi ekstraksi fitur. Teknik seperti *Gaussian blur* atau median filter digunakan untuk meredam *noise* tanpa menghilangkan detail penting pada wajah.

C. Pembagian Data: Data Latih dan Data Uji

Data latih digunakan untuk membangun model pembelajaran, sedangkan data uji digunakan untuk menguji sejauh mana model mampu mengenali pola baru, khususnya dalam membedakan antara santri yang memiliki indikasi perilaku merokok dan yang tidak. Kedua jenis data ini kemudian digunakan dalam penerapan dua algoritma utama, yaitu *Linear Discriminant Analysis* (LDA) dan *K-Nearest Neighbor* (K-NN). LDA digunakan untuk mengekstraksi dan mereduksi fitur dari citra wajah dengan cara memproyeksikan data ke ruang baru yang memaksimalkan separasi antar kelas. Setelah itu, K-NN digunakan untuk melakukan proses klasifikasi berdasarkan kedekatan jarak antara data uji dan data latih.

D. Ekstraksi Fitur

Ekstraksi fitur merupakan tahap penting dalam proses analisis data—yakni transformasi data mentah seperti citra, teks, atau sinyal menjadi representasi yang lebih ringkas dan padat informasi agar dapat dikelola lebih efektif oleh model klasifikasi atau klustering. Pada citra, fitur tekstur sering kali menjadi titik awal yang andal metode seperti *Gray Level Co-occurrence Matrix* (GLCM) dan *Local Binary Patterns* (LBP) terbukti efektif dalam menangkap pola permukaan; GLCM berguna untuk mengukur karakteristik seperti kontras, korelasi, energi, dan homogenitas, sementara LBP tangguh terhadap variasi iluminasi ringan dan mampu menyorot detail tekstur halus. Studi pada tanaman padi (Saniah et al., 2024) menggunakan GLCM dengan hasil yang menjanjikan dalam klasifikasi penyakit dan penelitian lain mengungkap bahwa kombinasi GLCM dan LBP memperkuat pengenalan pola tekstur dalam berbagai aplikasi.

E. Metode LDA

Pendekatan LDA, langkah-langkah yang dilakukan meliputi: Pelatihan LDA menggunakan data latih untuk mendapatkan fungsi diskriminan. Proyeksi LDA, yaitu mereduksi dimensi fitur sambil mempertahankan separasi antar kelas. Pengujian



dilakukan dengan data uji yang telah diproyeksikan, kemudian dihitung jaraknya ke pusat kelas. Hasilnya digunakan untuk melakukan klasifikasi, sama seperti pada KNN, tetapi dalam ruang hasil proyeksi. Data yang sudah diketahui labelnya digunakan untuk menemukan fungsi diskriminan. Untuk data yang belum diketahui kelasnya/labelnya, kita menggunakan fungsi diskriminan yang telah ditemukan.

F. Metode K-NN

Proses klasifikasi menggunakan metode *K-Nearest Neighbor* (K-NN) diawali dengan menentukan banyaknya tetangga terdekat yang akan digunakan, yaitu nilai K. Selanjutnya, dihitung jarak *Euclidean* antara fitur uji yang ingin diprediksi dengan semua fitur latih. Setelah itu, dipilih K objek fitur latih yang memiliki jarak terdekat dengan fitur uji tersebut. Terakhir, prediksi kelas dari fitur uji dilakukan berdasarkan K fitur latih terdekat yang telah dipilih.

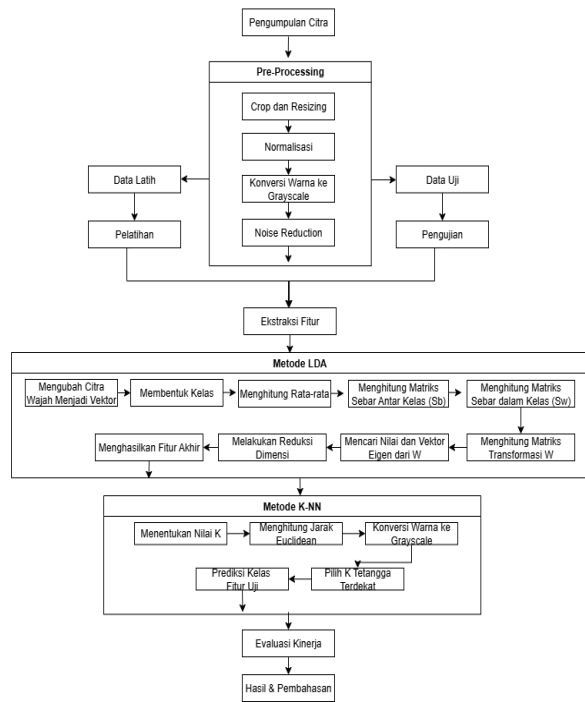
G. Evaluasi Kinerja

Evaluasi Kinerja merupakan proses sistematis yang digunakan untuk mengukur performa model pembelajaran mesin (machine learning) atau sistem klasifikasi berdasarkan metrik tertentu yang relevan dengan konteks data. Proses ini menjadi bagian krusial dalam pipeline pengembangan model karena bertujuan untuk mengetahui seberapa akurat dan andal model dalam memprediksi data baru serta untuk mengidentifikasi potensi kesalahan dan bias dalam sistem. Evaluasi kinerja sistem dengan menghitung metrik-metrik klasifikasi seperti: Akurasi, Presisi, *Recall*, dan *F1-Score*. Evaluasi ini penting untuk menilai efektivitas model K-NN dan LDA dalam mengklasifikasikan perilaku merokok secara otomatis dan menentukan metode mana yang lebih optimal.

H. Hasil dan Pembahasan

Hasil dari penelitian dianalisis berdasarkan pengujian data citra wajah santri baru yang telah diproses menggunakan algoritma *Linear Discriminant Analysis* (LDA) dan *K-Nearest Neighbor* (K-NN). Mekanisme kerja sistem dimulai dengan pengambilan citra wajah, kemudian dilakukan proses ekstraksi fitur dari citra tersebut. Fitur-fitur wajah yang telah diperoleh selanjutnya direduksi dimensinya menggunakan LDA, dengan tujuan memaksimalkan pemisahan antara dua kelas, yaitu santri perokok dan bukan perokok. Hasil dari LDA berupa data fitur yang lebih ringkas dan diskriminatif, kemudian digunakan sebagai input pada proses klasifikasi menggunakan algoritma K-

NN. K-NN bekerja dengan membandingkan fitur wajah yang diuji dengan data latih terdekat untuk menentukan kelas perilaku (perokok atau bukan perokok) berdasarkan nilai k tetangga terdekat.



Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

Hasil dari penelitian dianalisis berdasarkan pengujian data citra wajah santri baru yang telah diproses menggunakan algoritma *Linear Discriminant Analysis* (LDA) dan *K-Nearest Neighbor* (K-NN). Mekanisme kerja sistem dimulai dengan pengambilan citra wajah, kemudian dilakukan proses ekstraksi fitur dari citra tersebut. Fitur-fitur wajah yang telah diperoleh selanjutnya direduksi dimensinya menggunakan LDA, dengan tujuan memaksimalkan pemisahan antara dua kelas, yaitu santri perokok dan bukan perokok. Hasil dari LDA berupa data fitur yang lebih ringkas dan diskriminatif, kemudian digunakan sebagai input pada proses klasifikasi menggunakan algoritma K-NN. K-NN bekerja dengan membandingkan fitur wajah yang diuji dengan data latih terdekat untuk menentukan kelas perilaku (perokok atau bukan perokok) berdasarkan nilai k tetangga terdekat.



A. Dataset (Data Citra)

Penelitian ini peneliti membuat sebuah sistem untuk klasifikasi mengidentifikasi kategori



perokok dan tidak perokok pada santri. Adapun dalam penelitian kali ini menggunakan algoritma *K-Nearest Neighbor* mengklasifikasi data dan untuk menggunakan algoritma Linear Discriminant Analisis untuk mendapatkan sebaran dari dataset. Untuk mempermudah dalam penelitian kali ini, peneliti menggunakan pemrograman MATLAB R2023b untuk pembuatan sistem mengidentifikasi kategori perokok pada santri baru di Pesantren Kauman Muhammadiyah Padang Panjang. Peneliti menggunakan dua jenis data citra, yaitu citra wajah perokok dan citra wajah non-perokok. Jumlah total dataset yang digunakan sebanyak 50 citra, terdiri dari 20 citra wajah perokok dan 30 citra wajah santri. Untuk data latih, digunakan 70% citra yang terdiri dari 15 wajah perokok dan 20 wajah santri. Sementara itu, untuk data uji digunakan 30% citra, terdiri dari 5 wajah perokok dan 10 wajah santri. Seluruh citra dikumpulkan dalam format “.JPG” dengan resolusi 100x100 piksel. Setelah dataset terkumpul, langkah selanjutnya adalah mempersiapkan pelatihan. Pelatihan dan pengujian dilakukan menggunakan aplikasi MATLAB.

Tabel 1. Tabel Dataset

Nama Citra	Data Primer	Nama Citra	Data Sekunder
Citra_01		Data_01	





Peneliti menggunakan dua jenis data citra, yaitu citra wajah perokok dan citra wajah non-perokok. Jumlah total dataset yang digunakan sebanyak 50 citra, terdiri dari 30 citra wajah perokok dan 20 citra wajah santri. Untuk data latih, digunakan 70% citra yang terdiri dari 15 wajah perokok dan 20 wajah santri. Sementara itu, untuk data uji digunakan 30% citra, terdiri dari 10 wajah perokok dan 5 wajah santri. Seluruh citra dikumpulkan dalam format “.JPG” dengan resolusi 100x100 piksel.

B. Pre-Processing

Dataset yang sudah terkumpul, langkah selanjutnya adalah melakukan *pre-processing* dengan langkah pertama *Crop and Resizing* pada

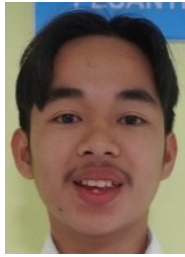

Tabel 2.

Tabel 2. Tabel *Crop and Resizing*

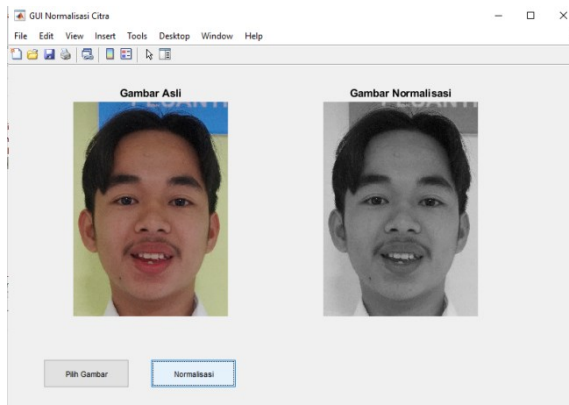
No	Data Primer		Data Sekunder	
	Citra_01	<i>Crop and Resizing</i>	Data_01	<i>Crop and Resizing</i>
1				

Tabel 2. menjelaskan proses *crop and resizing* pada citra bertujuan untuk memastikan citra tidak terlalu berlebih ukurannya. Selanjutnya, ukuran citra diseragamkan menjadi dimensi tetap, misalnya 100x100 piksel, agar konsisten pada seluruh data dan sesuai untuk proses klasifikasi. Langkah selanjutnya proses konversi warna ke dalam bentuk *grayscale*.

Tabel 3. Tabel Konversi Warna *Grayscale*

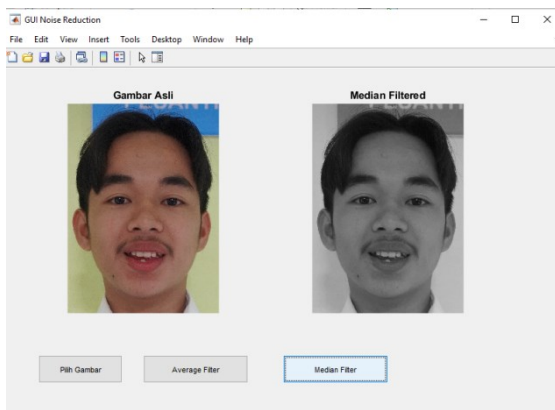
Nama Citra	Dataset Awal	Hasil <i>Grayscale</i>
Citra_01		

Tabel 3. citra wajah dikonversi ke skala abu-abu (*grayscale*) dengan tujuan menyederhanakan informasi visual dan mengurangi kompleksitas yang terkait dengan variasi warna. Proses konversi ini memungkinkan fokus pada intensitas piksel yang merepresentasikan struktur dan tekstur wajah, sehingga mempermudah tahap ekstraksi ciri. Selanjutnya, citra dinormalisasi sehingga nilai piksel berada dalam rentang tertentu, yang berfungsi untuk menstandarkan data input dan meningkatkan konsistensi serta performa algoritma klasifikasi pada tahap berikutnya.



Gambar 2. Normalisasi Citra

Gambar 2. menunjukkan bahwa proses normalisasi citra memiliki nilai piksel pada 34 citra asli yang biasanya berada pada rentang 0 hingga 255 (untuk citra 8-bit). Jika nilai piksel ini dibiarkan tanpa normalisasi, hal tersebut dapat menyebabkan ketidakstabilan pada proses pelatihan model LDA dan K-NN, karena perbedaan skala antar fitur akan memengaruhi perhitungan jarak dan koefisien diskriminan. Oleh karena itu, normalisasi dilakukan untuk menstandarkan rentang nilai piksel sehingga semua fitur memiliki kontribusi yang seimbang, meningkatkan konsistensi, stabilitas, dan akurasi klasifikasi citra wajah perokok maupun tidak perokok.

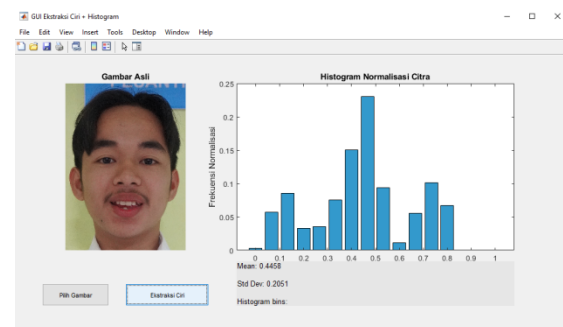


Gambar 3. Noise Reduction (Median Filter)

Proses pada Gambar 2. penting dalam pengolahan citra digital yang bertujuan untuk menghilangkan gangguan atau “noise” pada citra agar kualitas visual dan informasi citra meningkat. Noise biasanya muncul akibat faktor eksternal atau internal saat akuisisi citra, seperti gangguan sensor kamera, transmisi data yang buruk, atau kondisi pencahayaan yang tidak stabil.

C. Ekstraksi Fitur

Tahapan ini bertujuan untuk mengambil informasi atau ciri khas yang paling relevan dari data mentah, seperti gambar wajah, sehingga data tersebut dapat direpresentasikan dalam bentuk yang lebih sederhana namun tetap bermakna.



Gambar 4. Ekstraksi Fitur

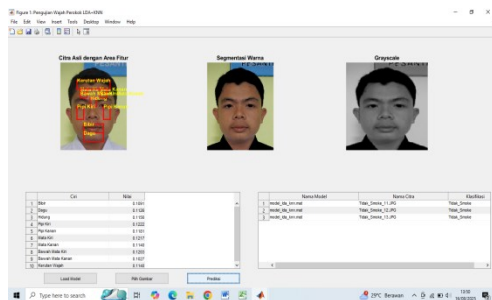
Gambar 4. menunjukkan proses ekstraksi ciri pada citra uji. Hasil ekstraksi menampilkan nilai mean sebesar 0,4458 dan standar deviasi (Std Dev) sebesar 0,2051, yang menggambarkan karakteristik intensitas piksel citra. Selain itu, tampilan histogram bins memperlihatkan distribusi frekuensi nilai piksel, sehingga memudahkan analisis variasi dan kontras pada citra.

D. Hasil LDA dan K-NN

Pada Tabel 4. disajikan hasil ekstraksi ciri dari seluruh data uji yang dianalisis menggunakan metode *Linear Discriminant Analysis* (LDA) dan *K-Nearest Neighbor* (K-NN). Ekstraksi ciri ini menghasilkan vektor fitur yang merepresentasikan karakteristik penting setiap citra, yang selanjutnya digunakan untuk proses klasifikasi. Hasil ini memungkinkan evaluasi performa model serta interpretasi kontribusi masing-masing fitur terhadap akurasi prediksi citra uji. Gambar berikut merupakan contoh penerapan pengolahan citra digital yang menggabungkan berbagai metode canggih, mulai dari segmentasi warna hingga klasifikasi berbasis pembelajaran mesin. Analisis ini tidak hanya membantu mengidentifikasi fitur-fitur spesifik dalam citra, tetapi juga mampu membedakan kategori objek dengan presisi tinggi. Gambar 5. dapat melihat secara detail bagaimana suatu citra diproses melalui tahapan-tahapan teknis, mulai dari ekstraksi fitur hingga hasil klasifikasi akhir. Hasil yang ditampilkan mencerminkan integrasi antara teknologi pengolahan citra dan kecerdasan buatan, yang memiliki potensi besar dalam berbagai aplikasi, seperti pemantauan lingkungan, sistem keamanan, atau deteksi dini



kebakaran.



Gambar 5. Normalisasi Citra

Gambar 5. menampilkan citra asli yang telah dianalisis menggunakan berbagai teknik pengolahan citra untuk mengidentifikasi fitur-fitur tertentu. Pada bagian kiri, terdapat beberapa area fitur yang diberi label seperti Bibir, Dagu, Hidung, Pipi kiri, pipi kanan, mata kiri, mata kanan, bawah mata kiri, bawah Mata kanan, Kerutan Wajah, yang kemungkinan merujuk pada bagian atau karakteristik spesifik dari citra. Sementara itu, bagian kanan menampilkan daftar fitur seperti Ciri dan Niki, disertai nilai numerik dari 0.1091 hingga 10, yang mungkin merupakan hasil pengukuran atau parameter terkait fitur tersebut. Proses segmentasi warna dilakukan menggunakan model yang tersimpan dalam file model_Ida_Jam.mat, bertujuan untuk memisahkan citra berdasarkan komponen warnanya. Selain itu, citra juga dikonversi ke dalam bentuk grayscale dengan contoh file seperti Tidak_Smoke_11.JPG, Tidak_Smoke_12.JPG, dan Tidak_Smoke_13.JPG, untuk mempermudah analisis tekstur atau intensitas cahaya.

Proses pada Gambar 4. penting dalam pengolahan citra digital yang bertujuan untuk menghilangkan gangguan atau “noise” pada citra agar kualitas visual dan informasi citra meningkat. Noise biasanya muncul akibat faktor eksternal atau internal saat akuisisi citra, seperti gangguan sensor kamera, transmisi data yang buruk, atau kondisi pencahayaan yang tidak stabil.

Tabel 4. Hasil Akurasi LDA dan K-NN

Nama Citra Uji	Citra Asli	Hasil Identifikasi	Confidence	Benar (Y/N)
Tidak_Smoke_11		Tidak_Smoke	100%	Y
Tidak_Smoke_12		Tidak_Smoke	99%	Y
Tidak_Smoke_13		Tidak_Smoke	100%	Y
Tidak_Smoke_14		Tidak_Smoke	100%	Y
Tidak_Smoke_15		Tidak_Smoke	100%	Y
Tidak_Smoke_16		Smoke	53%	N
Tidak_Smoke_19		Tidak_Smoke	100%	Y
Tidak_Smoke_20		Tidak_Smoke	100%	Y

Berdasarkan hasil pengujian, sebagian besar data uji dengan label Tidak_Smoke berhasil diidentifikasi dengan benar, dengan tingkat confidence yang sangat tinggi, berkisar antara 99% hingga 100%. Meskipun demikian, terdapat satu kasus kesalahan klasifikasi pada citra



Tidak_Smoke2_16, di mana sistem memprediksi kelas Smoke dengan tingkat confidence sebesar 41%, yang tidak sesuai dengan label aslinya. Temuan ini mengindikasikan bahwa sistem memiliki performa klasifikasi yang tinggi, namun masih terdapat potensi kesalahan, khususnya pada citra yang memiliki karakteristik visual serupa antar kelas. Secara umum, kinerja sistem dapat dikategorikan baik, mengingat tingkat akurasi yang dicapai mendekati sempurna pada sebagian besar data uji. Untuk mengukur kinerja sistem identifikasi citra wajah perokok dan tidak perokok, digunakan metrik akurasi. Metrik ini menunjukkan persentase jumlah prediksi yang benar dibandingkan dengan seluruh data uji yang digunakan. Adapun perumusan akurasi dapat dinyatakan sebagai berikut:

$$\text{Akurasi}(\%) = \frac{\text{Jumlah Prediksi Benar}}{\text{Jumlah Seluruh Data Uji}} \times 100\% \quad (1)$$

Rumus (1) digunakan untuk mendapatkan tingkat akurasi sebagai berikut:

$$\text{Akurasi}(\%) = \frac{7}{8} \times 100\% = 87\%$$

Akurasi sistem pada pengujian ini adalah 87,5%, yang menunjukkan bahwa sistem mampu mengklasifikasikan citra wajah perokok dan tidak perokok dengan ketepatan tinggi, meskipun masih terdapat satu kesalahan klasifikasi pada citra dengan tingkat *confidence* 75%.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem identifikasi kategori perokok pada citra wajah santri dengan memanfaatkan kombinasi algoritma Linear Discriminant Analysis (LDA) dan K-Nearest Neighbor (K-NN). Sistem ini mencapai akurasi sebesar 87,5% dalam membedakan santri perokok dan non-perokok, yang menunjukkan efektivitasnya dalam analisis citra wajah. Proses diawali dengan pre-processing citra, termasuk cropping, resizing, konversi grayscale, normalisasi, dan noise reduction, untuk memastikan kualitas input yang konsisten. Selanjutnya, ekstraksi fitur menggunakan LDA berperan penting dalam mereduksi dimensi data sekaligus memaksimalkan pemisahan antar kelas, sementara K-NN digunakan untuk klasifikasi berdasarkan kedekatan jarak fitur. Meskipun demikian, sistem masih mengalami satu kesalahan klasifikasi, di mana citra non-perokok teridentifikasi sebagai perokok dengan tingkat *confidence* 53%. Hal ini mengindikasikan adanya keterbatasan dalam menangani kemiripan visual tertentu, yang mungkin dipengaruhi oleh faktor

pencahayaan atau ekspresi wajah. Selain itu, ukuran dataset yang relatif kecil (50 citra) dan ketidakseimbangan jumlah sampel (30 citra perokok vs. 20 citra non-perokok) turut memengaruhi kinerja model.

Kontribusi penelitian ini terletak pada integrasi LDA dan K-NN yang terbukti mampu meningkatkan akurasi identifikasi, serta potensi aplikasinya dalam lingkungan pesantren sebagai alat pendeteksi perilaku menyimpang berbasis citra. Hasil penelitian juga menjadi landasan untuk pengembangan lebih lanjut, seperti perluasan dataset dengan variasi yang lebih beragam, eksperimen menggunakan algoritma lain seperti Convolutional Neural Network (CNN), atau implementasi sistem real-time berbasis CCTV. Rekomendasi untuk penelitian selanjutnya mencakup peningkatan kuantitas dan kualitas data, eksplorasi fitur tambahan seperti tekstur kulit atau noda nikotin, serta pengujian dalam skenario real-time untuk memvalidasi keandalan sistem di kondisi nyata. Dengan demikian, penelitian ini tidak hanya memberikan solusi teknologi untuk masalah kesehatan di pesantren, tetapi juga membuka peluang pengembangan sistem identifikasi berbasis machine learning yang lebih canggih dan adaptif di masa depan.

5. REFERENSI

- Akbar, F. M. N. (2024). Metode KNN (K-Nearest Neighbor) untuk menentukan kualitas air. *Jurnal Tekno Kompak*, 18(1), 28–40. <https://doi.org/10.XXXX/jtk.2024.18.1.28>
- Alim, S. (2024). *Ulama dan kolonialisme Belanda: Respons Syekh Nawawi Banten dan Sayid Usman*. Thalibul Ilmi Publishing & Education.
- Dwinda, T. (2025). Hubungan antara kecemasan dengan perilaku merokok pada remaja siswa SMK Negeri 1 Pulau Punung. <https://doi.org/10.XXXX/jurnal2025.970>
- Febrianto, M. R., Wijaya, N. K., & Rilvani, E. (2025). Penerapan algoritma C4.5 untuk prediksi kelulusan mahasiswa berdasarkan data akademik. *Jurnal Media Akademik (JMA)*, 3(8). <https://doi.org/10.XXXX/jma.2025.3.8>
- Gunarsih, N. A., Wiyono, S., & Afidah, D. I. (2024). Aplikasi deteksi pelanggaran merokok di tempat larangan merokok. *Jurnal Informatika Polinema*, 10(3), 351–358. <https://doi.org/10.XXXX/jip.2024.10.3.351>



- Haikhah, J. (2021). Persepsi santri perokok terhadap anjuran kiai untuk tidak merokok di Pondok Pesantren Sunan Pandanaran Komplek IV Yogyakarta.
<https://doi.org/10.XXXX/jurnal2021.001>
- Isnaini, S. N. R. (2023). Efektivitas bimbingan agama dalam mengurangi perilaku agresif anak berhadapan dengan hukum (ABH) di Sentra Mulya Jaya Jakarta.
<https://doi.org/10.XXXX/jurnal2023.002>
- Karim, A. (2025). Big data analytics: Analisis sentimen netizen di era media baru. Penerbit NEM.
- Khaliqah, M., Sarifah, L., & Khotijah, S. (2024). Implementasi algoritma K-Nearest Neighbor (K-NN) dalam mengklasifikasikan berbagai jenis ekspresi wajah manusia. *Zeta-Math Journal*, 9(1), 10–20.
<https://doi.org/10.XXXX/zmj.2024.9.1.10>
- Khirusisah, F. (2024). Perancangan program konseling Islam dalam mengatasi perilaku menyimpang santri di Pondok Pesantren Safinatussalamah Gampong Pintu Gayo Kecamatan Putri Betung Kabupaten Gayo Lues.
<https://doi.org/10.XXXX/jurnal2024.003>
- Lubis, S., Yuningsih, Y., Marbun, R. A., Tarigan, S., & Achyar, A. J. (2025). Analisis faktor-faktor yang mempengaruhi kenakalan remaja di kalangan pelajar SMK Negeri 1 Percut Sei Tuan. *Ranah Research: Journal of Multidisciplinary Research and Development*, 7(3), 1480–1494.
<https://doi.org/10.XXXX/ranah.2025.7.3.1480>
- Maulani, G., Hasan, F. N., Setiawan, D., Bowo, I. T., Ardhana, V. Y. P., Ramdhani, Y., Inayah, I., et al. (2025). *Machine Learning*. Mega Press Nusantara.
- Meldyantono, A. P. (2025). Implementasi sistem absensi berbasis pengenalan wajah menggunakan metode CNN dan model FaceNet.
<https://doi.org/10.XXXX/jurnal2025.004>
- Muflikhah, L., Mahmudy, W. F., & Kurnianingtyas, D. (2023). *Machine learning*. Universitas Brawijaya Press.
- Nurdiansyah, A., Erlanda, H., Syafril, S., Roza, Y. B., & Sovia, R. (2025). Klasifikasi citra dalam identifikasi kol dan wortel menggunakan algoritma LDA dan KNN. *Journal of Science and Social Research*, 8(2), 1895–1902.
<https://doi.org/10.XXXX/jssr.2025.8.2.1895>
- Nurlina, N. (2024). Perilaku merokok masyarakat dan dampak sosialnya terhadap remaja di Kecamatan Suppa Kabupaten Pinrang.
<https://doi.org/10.XXXX/jurnal2024.005>
- Rahmawati, D., Wahyuni, E. D., & Kartika, D. S. Y. (2025). Analisis sentimen berbasis aspek pada respons survei open-ended menggunakan LDA dan SVM. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4628–4634.
<https://doi.org/10.XXXX/jati.2025.9.3.4628>
- Rifai, M. H., Pramudya, D. A., & Narfandi, R. R. (2024). Analisis peran teknologi kecerdasan buatan dalam mengoptimalkan proses deteksi terhadap serangan siber. In *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis* (pp. 495–502).
<https://doi.org/10.XXXX/prosiding2024.495>
- Vladea Rahman, A. (2023). Hubungan tingkat pengetahuan dan dukungan keluarga dengan kepatuhan diet dan kontrol tekanan darah pada penderita hipertensi.
<https://doi.org/10.XXXX/jurnal2023.006>
- Yusmah Sari, J., & Ningrum, I. P. (2017). Pengenalan wajah menggunakan metode Linear Discriminant Analysis dan k-Nearest Neighbor. *ULTIMATICS*, IX(1).
<https://doi.org/10.XXXX/ultimatics.2017.IX.1>