

Laporan Pembuatan Sistem Identifikasi Jenis Kelamin Otomatis Berdasarkan Geometri Wajah Menggunakan OpenCV dan Python



Dosen Pengampuh:

Ali Rohman, M.Kom.

Disusun Oleh:

- 1. Yunita Rachmawati (14012200165)**
- 2. Cinta Azzahra Z (14012200164)**
- 3. Adelia Dara Puspita (14012200222)**
- 4. Deva Delyana Ruanda (14012200120)**
- 5. Uroh Humairoh (14012200129)**

**PRODI ILMU KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BINA BANGSA**

BAB I

PENDAHULUAN

A. Latar Belakang

Di era digital yang semakin pesat berkembang ini, pemanfaatan teknologi pengenalan wajah (face recognition) telah merambah ke berbagai bidang. Salah satu aspek menarik dalam pengenalan wajah adalah kemampuan system untuk mendeteksi atau mengklasifikasi jenis kelamin seseorang berdasarkan fitur – fitur geometri wajah. Deteksi jenis kelamin otomatis ini dapat meningkatkan efisiensi dalam system interaktif manusia mesin.

Geometri wajah mencakup struktur dan proses elemen – elemen wajah seperti jarak antara mata, panjang hidung, lebar rahang, dan bentuk dagu. Fitur – fitur ini secara statistic memiliki perbedaan antara wajah laki – laki dan perempuan, yang dapat dianalisis menggunakan pendekatan pembelajaran mesin (mechine learning) atau pembelajaran mendalam (deep learning). Dengan memanfaatkan teknologi tersebut, system dapat dilatih untuk mengenali pola – pola tertentu yang menjadi indicator jenis kelamin.

Python merupakan salah satu bahasa pemrograman yang populer dalam pengembangan system berbasis kecerdasan buatan karena ekosistem pustaka (library) yang kaya seperti OpenCV, Dlib, dan scikit-learn. PyCharm sebagai lingkungan pengembangan terintegrasi (IDE) memberikan kemudahan bagi pengembang dalam menulis, menguji, dan menjalankan kode Python dengan efisien.

Oleh karna itu, pengembangan system ini bertujuan untuk membangun sebuah system pendeteksi jenis kelamin otomatis berbasis geometri wajah yang dibangun menggunakan bahasa pemrograman Python di dalam lingkungan pengembangan PyCharm. Sistem ini diharapkan dapat mendeteksi jenis kelamin seseorang secara akurat dalam gambar wajah dengan memanfaatkan fitur geometris yang diekstraksi melalui proses deteksi dan pelacakan titik – titik wajah (facial landmarks).

B. Tujuan

Tujuan dari pengembangan system ini adalah sebagai berikut :

1. Membangun system yang dapat mendeteksi wajah dan citra input menggunakan Python.
2. Mengekstraksi fitur geometri wajah melalui deteksi titik – titik wajah (facial landmarks).
3. Mengembangkan model klasifikasi untuk membedakan jenis kelamin berdasarkan fitur – fitur geometris wajah.
4. Menerapkan dan menguji system secara keseluruhan dalam lingkungan PyCharm untuk memastikan fungsionalitas dan akurasi system.

C. Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah yang dapat dirumuskan adalah:

1. Bagaimana cara mendeteksi wajah dan mengekstraksi fitur geometris dari citra wajah menggunakan Python ?
2. Fitur geometris apa saja yang relevan untuk membedakan jenis kelamin wajah secara akurat ?
3. Metode klasifikasi apa yang paling efektif untuk membedakan jenis kelamin berdasarkan fitur wajah ?
4. Seberapa akurat system pendeteksi jenis kelamin yang dibangun dengan pendekatan ini dalam pengujian nyata ?

BAB II

LANDASAN TEORI

A. Pengolahan Citra Digital

Pengolahan citra digital merupakan proses manipulasi citra menggunakan computer untuk meningkatkan kualitas visual mengekstrak informasi tertentu. Dalam konteks system ini, pengolahan citra digunakan untuk mendeteksi wajah, mengekstraksi fitur, dan melakukan klasifikasi.

B. Pengenalan Sisten Pendeteksi Wajah (Face Detection)

Deteksi wajah merupakan langkah awal dalam banyak aplikasi pengolahan citra seperti pengenalan wajah, identifikasi emosi, dan klasifikasi jenis kelamin. Salah satu metode populer yang digunakan adalah Haar Cascade Classifier, algoritma berbasis fitur Haar yang dikembangkan oleh Viola dan Jones. Metode ini bekerja dengan mendeteksi pola-pola khusus pada wajah manusia berdasarkan gradasi intensitas piksel, OpenVC dan Dlib.

C. Facial Landmark Detection

Facial landmarks adalah titik-titik kunci pada wajah seperti ujung mata, hidung, mulut, dan tepi wajah. Dlib menyediakan model pre-trained yang mendeteksi 68 titik pada wajah yang digunakan untuk mengekstrak fitur geometris seperti jarak antar titik dan sudut.

D. Ekstraksi Fitur Geometri

Dalam deteksi berbasis geometri, fitur utama yang digunakan berasal dari bentuk dan ukuran wajah, seperti lebar hidung, jarak antar mata, dan bentuk rahang. Dengan menggunakan metode deteksi titik landmark pada wajah, seperti 68 facial landmarks, sistem dapat mengekstrak fitur-fitur numerik yang dapat digunakan untuk klasifikasi.

E. Klasifikasi Jenis Kelamin

Setelah fitur wajah diekstraksi, proses klasifikasi dilakukan menggunakan algoritma pembelajaran mesin. Salah satu model yang efektif dan ringan untuk tugas ini adalah Support Vector Machine (SVM), yang dapat memisahkan dua kelas (laki-laki dan perempuan) dengan garis keputusan terbaik pada ruang fitur berdimensi tinggi. Model ini dilatih menggunakan fitur geometris dari wajah yang telah dikumpulkan sebelumnya.

F. Antarmuka Pengguna

Antarmuka pengguna dibangun menggunakan framework Streamlit, sebuah pustaka Python yang memudahkan pembuatan aplikasi web interaktif untuk keperluan data science dan machine learning. Dengan Streamlit, pengguna dapat mengunggah gambar, mengambil foto, dan langsung melihat hasil prediksi dari model yang digunakan.

G. Geometri Wajah dalam Pengenalan Jenis Kelamin

Geometri wajah merujuk pada hubungan spasial antara titik-titik karakteristik wajah. Penelitian menunjukkan bahwa proporsi dan dimensi wajah laki-laki dan perempuan berbeda secara statistik. Misalnya, laki-laki cenderung memiliki rahang lebih lebar dan hidung lebih panjang, sementara perempuan memiliki tulang pipi menonjol dan dagu meruncing. Representasi fitur-fitur ini sebagai jarak antar titik landmark menjadi dasar pembentukan vektor numerik yang berguna untuk klasifikasi.

H. Support Vector Machine (SVM)

SVM merupakan algoritma klasifikasi yang kuat untuk data berdimensi tinggi. Dalam klasifikasi jenis kelamin, SVM mencari hyperplane terbaik untuk memisahkan dua kelas (laki-laki dan perempuan). Keunggulan SVM:

- Cocok untuk dataset kecil hingga menengah.
- Memiliki margin maksimum yang optimal untuk pemisahan kelas.
- Dapat ditingkatkan dengan kernel non-linear seperti RBF dan polynomial.

I. Pickle: Serialisasi Model

Pickle adalah pustaka Python untuk menyimpan dan memuat objek Python dalam format biner. Dengan menggunakan Pickle, model hasil pelatihan dapat disimpan dan dimuat kembali saat aplikasi dijalankan tanpa perlu pelatihan ulang. Ini berguna dalam proses deployment model ke antarmuka aplikasi.

J. Streamlit sebagai Framework Antarmuka

Streamlit memungkinkan pembuatan antarmuka berbasis web yang sederhana dan efisien untuk proyek machine learning. Tanpa perlu HTML atau JavaScript, pengembang dapat membangun dashboard dan sistem interaktif hanya dengan script Python. Fitur utamanya meliputi:

- Upload file gambar
- Integrasi kamera
- Tampilan hasil klasifikasi langsung
- Navigasi antarmuka berbasis sidebar dan tombol

K. Evaluasi Model Machine Learning

Evaluasi model dilakukan untuk mengukur performa prediksi terhadap data yang belum pernah dilihat. Metrik evaluasi umum:

- **Accuracy:** persentase prediksi yang benar.
- **Precision:** proporsi benar dari seluruh prediksi positif.
- **Recall:** proporsi kasus positif yang berhasil diprediksi.
- **F1-Score:** harmonisasi antara precision dan recall.

Metrik-metrik ini dapat dihitung menggunakan fungsi `classification_report` dari pustaka `scikit-learn`.

L. Deep Learning dalam Klasifikasi Wajah

Selain metode berbasis fitur geometri, pendekatan **deep learning** telah banyak digunakan dalam pengenalan wajah dan klasifikasi jenis kelamin. Deep learning memungkinkan model untuk belajar langsung dari citra mentah tanpa perlu ekstraksi fitur manual.

Salah satu arsitektur yang populer adalah **Convolutional Neural Network (CNN)**, yang memiliki keunggulan dalam mengenali pola spasial pada gambar. CNN bekerja dengan menerapkan filter konvolusi yang secara otomatis mendeteksi fitur penting seperti tepi, bentuk mata, atau hidung. Framework umum yang digunakan untuk membangun CNN adalah **TensorFlow** dan **PyTorch**.

M. Augmentasi Data

Untuk mengatasi keterbatasan jumlah data latih, digunakan teknik **augmentasi data**, yaitu membuat variasi gambar baru dari data yang sudah ada. Contohnya termasuk:

- Rotasi gambar
- Perubahan pencahayaan
- Pembesaran (zoom)
- Flip horizontal
- Crop sebagian gambar

Teknik ini membantu model menjadi lebih tangguh terhadap variasi pose, ekspresi, dan kondisi pencahayaan pada wajah.

N. Normalisasi dan Standarisasi Fitur

Sebelum fitur digunakan dalam pelatihan model, penting untuk melakukan **normalisasi** atau **standarisasi**. Ini bertujuan agar semua fitur berada pada skala yang sama dan model tidak bias terhadap fitur dengan rentang nilai lebih besar. Teknik umum mencakup:

- **Min-Max Scaling**
- **Z-Score Normalization**

O. Cross-Validation

Cross-validation digunakan untuk menguji generalisasi model terhadap data yang tidak terlihat. Teknik populer seperti **K-Fold Cross Validation** membagi data menjadi beberapa subset (folds), kemudian model dilatih dan diuji secara bergiliran. Hasilnya dirata-rata untuk mendapatkan estimasi performa model yang lebih stabil dan akurat

BAB III

METODOLOGI

A. Studi Literatur

Tahapan awal melibatkan pencarian dan pemahaman terhadap literatur yang berkaitan dengan:

- Deteksi wajah menggunakan OpenCV
- Ekstraksi ciri geometri wajah
- Klasifikasi jenis kelamin dengan machine learning
- Implementasi antarmuka menggunakan Streamlit

Pada tahap ini dilakukan identifikasi terhadap kebutuhan sistem, baik dari sisi teknis maupun fungsional, yaitu:

- Sistem mampu mendeteksi wajah dari citra atau kamera.
- Sistem dapat mengekstrak fitur geometris wajah sebagai input model.
- Sistem dapat mengklasifikasikan jenis kelamin berdasarkan fitur wajah.
- Sistem menyimpan riwayat deteksi.
- Sistem memiliki autentikasi login dan registrasi pengguna.

B. Pengumpulan dan Persiapan Dataset

Dataset digunakan untuk pelatihan model klasifikasi. Dataset tersebut berisi gambar wajah laki-laki dan perempuan, dengan label yang telah ditentukan. Setiap gambar diproses untuk mengekstrak fitur wajah berdasarkan bentuk dan struktur geometris.

Sumber data yang digunakan berasal dari dataset wajah publik seperti:

- UTKFace Dataset
- Adience Benchmark

Langkah-langkah pada tahap ini:

- Mengunduh dataset wajah beranotasi jenis kelamin.
- Melakukan labeling jika dataset belum lengkap.
- Melakukan pemrosesan awal seperti cropping wajah, konversi skala abu-abu, dan normalisasi ukuran (misal 100x100 piksel).

C. Pra-pemrosesan

Gambar yang dimasukkan akan dikonversi menjadi skala abu-abu, kemudian dilakukan deteksi wajah menggunakan Haar Cascade. Setelah wajah terdeteksi, wajah dipotong dan diubah ukurannya menjadi resolusi tetap (misal: 100x100 piksel) untuk diekstraksi sebagai vektor fitur.

D. Ekstraksi Fitur Wajah

Untuk setiap gambar wajah, dilakukan deteksi dan ekstraksi fitur geometri. Proses ini melibatkan:

- Deteksi wajah menggunakan Haar Cascade.
- (Opsional) Deteksi facial landmarks (68 titik fitur wajah) jika digunakan.
- Flattening citra wajah menjadi vektor fitur satu dimensi (misal 10.000 dimensi dari 100x100 piksel).

Output dari tahap ini adalah kumpulan data fitur wajah dan label jenis kelamin untuk proses pelatihan model.

E. Pelatihan Model

Model SVM dilatih menggunakan fitur wajah yang telah diproses sebelumnya. Label jenis kelamin digunakan untuk membimbing proses pembelajaran. Setelah pelatihan, model disimpan menggunakan pickle untuk digunakan dalam aplikasi Streamlit.

Setelah fitur diperoleh, model klasifikasi jenis kelamin dilatih menggunakan algoritma Support Vector Machine (SVM).

Langkah-langkah pelatihan:

- Split data ke dalam data latih dan data uji (misalnya 80:20).
- Melatih model menggunakan SVC(probability=True) dari pustaka sklearn.
- Menyimpan model terlatih ke dalam file .pkl menggunakan pickle untuk digunakan kembali tanpa melatih ulang.

F. Pembuatan Antarmuka Aplikasi

Pengembangan antarmuka pengguna dilakukan dengan menggunakan Streamlit, yang mencakup:

- Halaman Login & Registrasi: menyimpan data pengguna dalam file CSV (users.csv).
- Halaman Beranda: memberikan petunjuk penggunaan.
- Halaman Upload Gambar / Kamera: menerima input dari pengguna dan melakukan prediksi jenis kelamin.
- Halaman Riwayat: menampilkan hasil deteksi terdahulu berdasarkan akun pengguna.

G. Implementasi Penyimpanan Riwayat

Setiap prediksi yang dilakukan disimpan dalam file riwayat.csv dengan informasi:

- Timestamp prediksi
- Hasil jenis kelamin
- Akurasi prediksi
- Username pengguna

Fitur ini berguna untuk analisis dan tracking histori penggunaan sistem oleh masing-masing pengguna.

H. Pengujian Sistem

Sistem diuji menggunakan gambar yang belum pernah dilihat oleh model (data uji). Evaluasi dilakukan berdasarkan tingkat keberhasilan model dalam memprediksi jenis kelamin dan stabilitas aplikasi saat digunakan oleh banyak pengguna.

Sistem diuji untuk memverifikasi bahwa semua fitur berjalan sesuai dengan tujuan:

- Uji Fungsi Deteksi: apakah wajah dapat dikenali dengan baik dari berbagai sudut/cahaya.
- Uji Akurasi Model: dengan menghitung akurasi prediksi pada data uji.
- Uji Antarmuka: memastikan bahwa login, upload, deteksi, dan riwayat berjalan lancar.

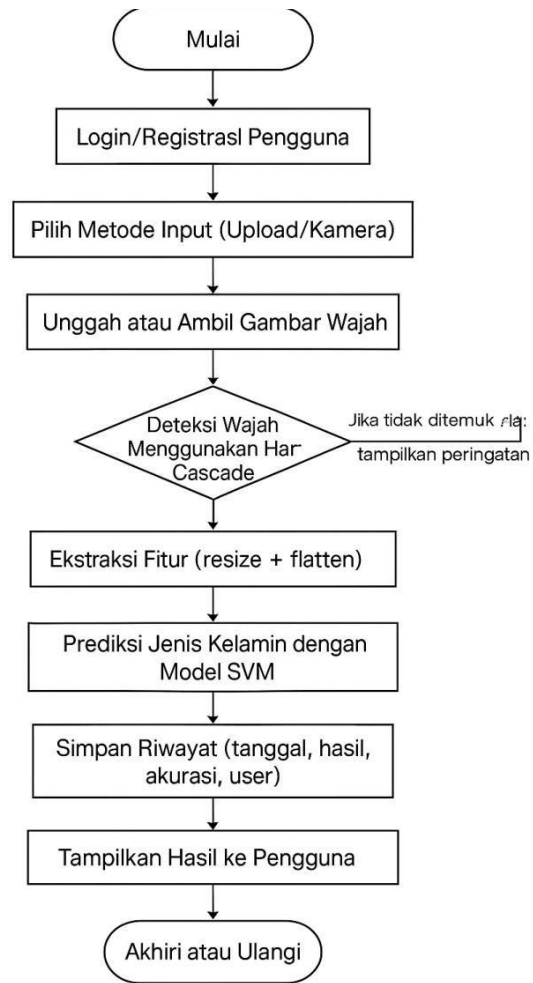
Jenis pengujian meliputi:

- Black Box Testing (pengujian fungsionalitas)
- Performance Testing (pengujian waktu proses prediksi)
- Usability Testing (respon pengguna terhadap antarmuka)

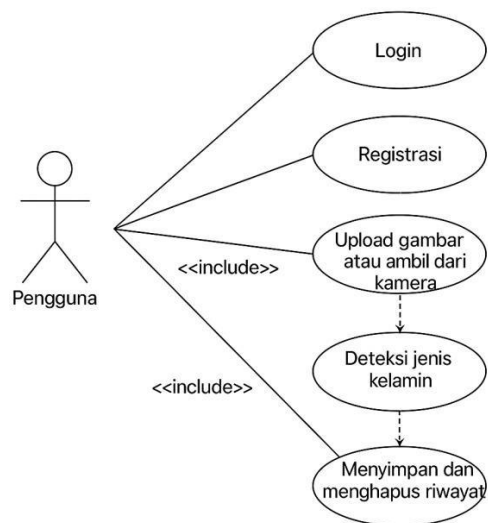
I. Diagram Alur Sistem

Berikut adalah diagram alir umum untuk sistem deteksi jenis kelamin berbasis geometri wajah:

1. Mulai
2. Login/Registrasi Pengguna
3. Pilih Metode Input (Upload/Kamera)
4. Unggah atau Ambil Gambar Wajah
5. Deteksi Wajah Menggunakan Haar Cascade
 - Jika tidak ditemukan wajah: tampilkan peringatan
6. Ekstraksi Fitur (resize + flatten)
7. Prediksi Jenis Kelamin dengan Model SVM
8. Simpan Riwayat (tanggal, hasil, akurasi, user)
9. Tampilkan Hasil ke Pengguna
10. Akhiri atau Ulangi



J. Use Case Diagram



K. Mockup Antarmuka

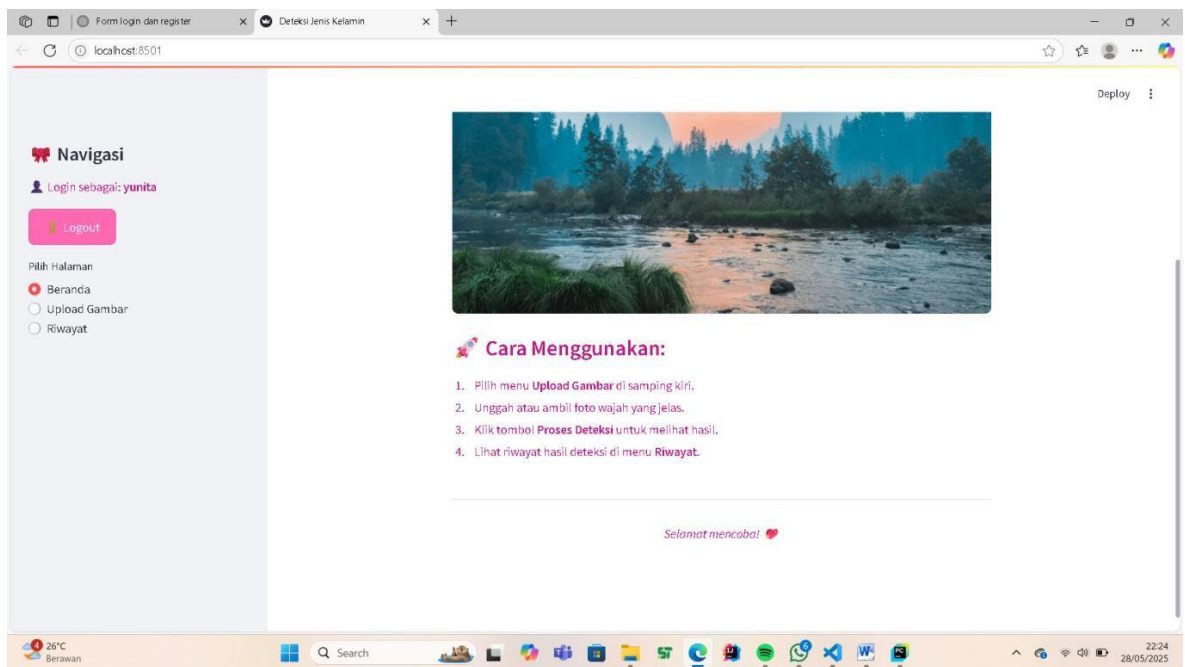
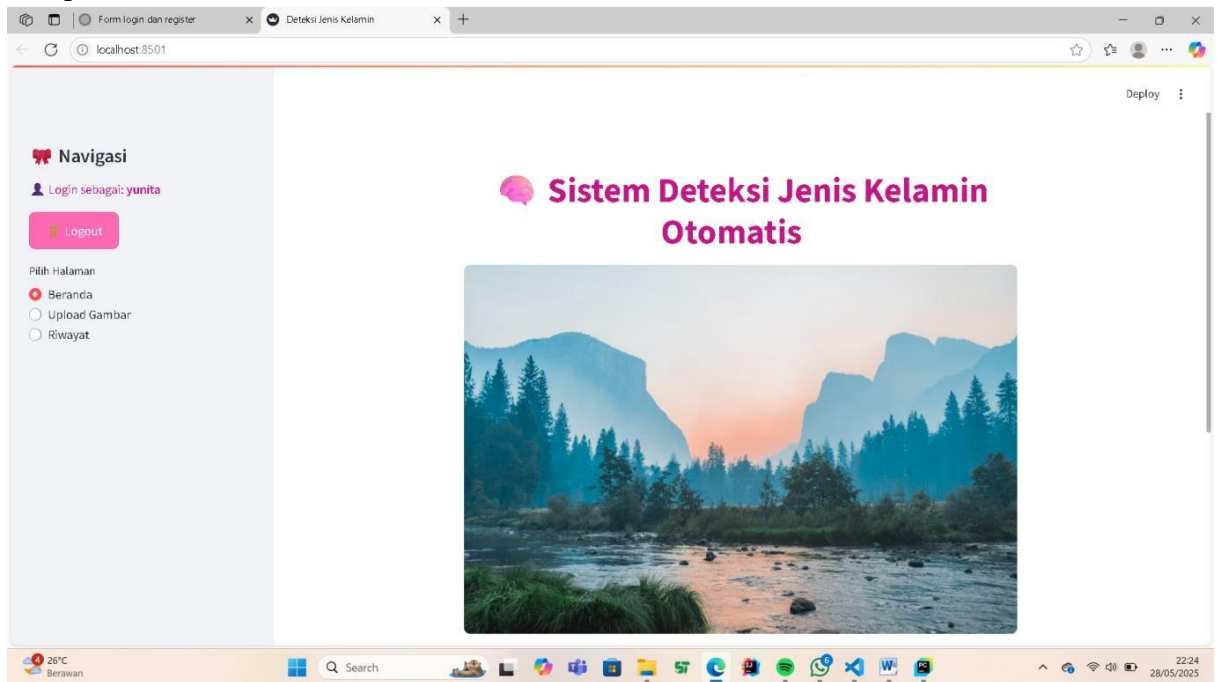
Tampilan registrasi,

The screenshot shows a web browser window with two tabs: 'Form login dan register' and 'Login'. The address bar shows 'localhost:8501'. On the left, a sidebar menu contains 'Menu', 'Login', and 'Registrasi' (highlighted with a red circle). The main content area has a header 'Sistem Deteksi Jenis Kelamin' with a speech bubble icon. Below it, a message says 'Silakan login atau registrasi untuk melanjutkan'. A large pink button labeled 'Registrasi' with a plus icon is centered. Below this, the text 'Buat akun baru' is displayed. The registration form includes three input fields: 'Username', 'Email', and 'Password' (with a toggle eye icon). A pink 'Registrasi' button with a checkmark icon is at the bottom of the form. The Windows taskbar at the bottom shows the date and time as 22:23 on 28/05/2025.

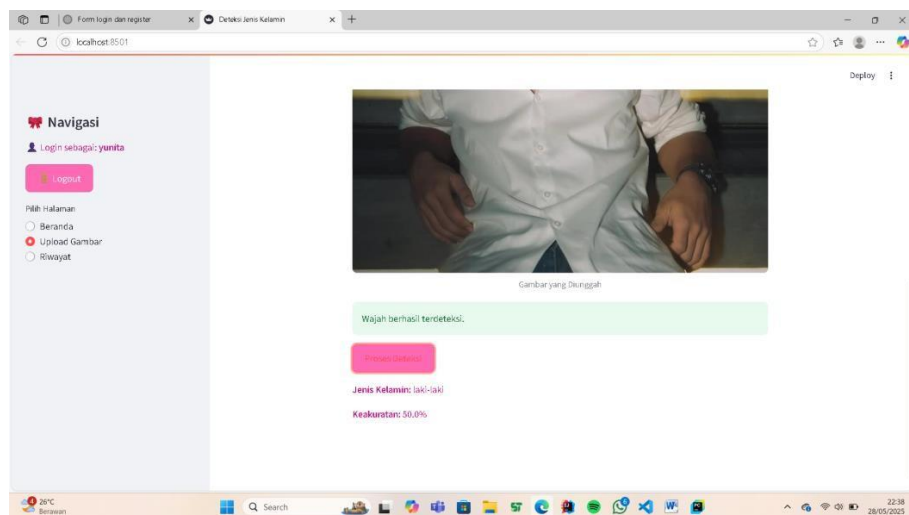
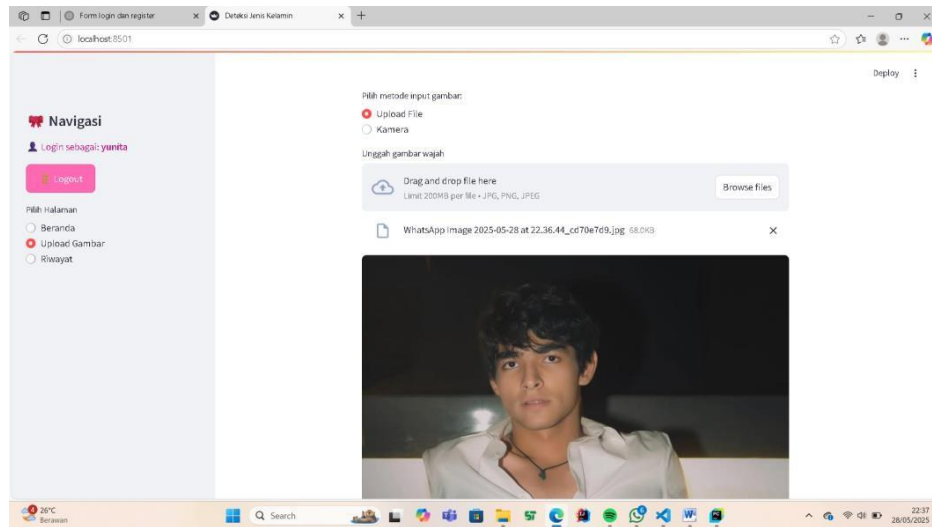
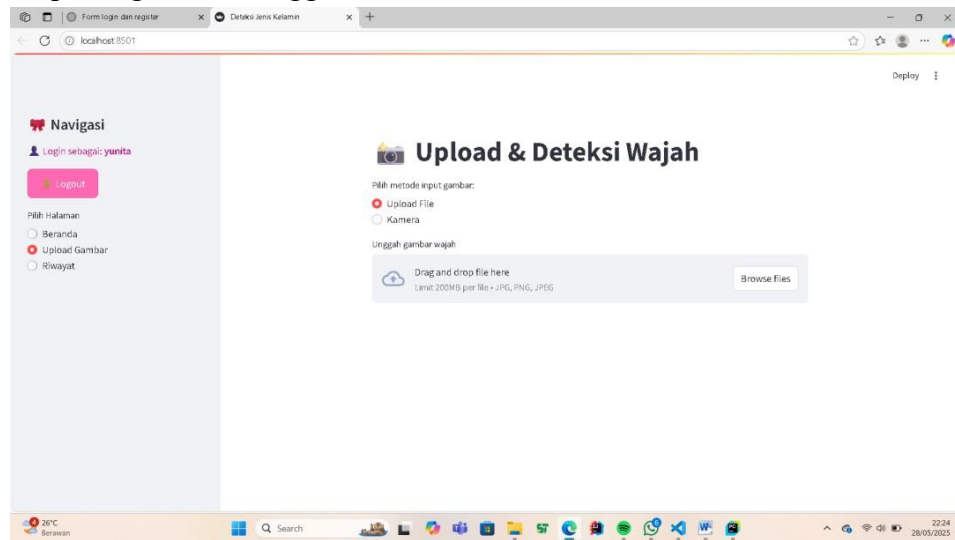
From login,

The screenshot shows the same web browser window, but the 'Login' tab is active. The sidebar menu remains the same. The main content area features the same header and message. A large pink button labeled 'Login' with a key icon is centered. Below it, the text 'Masukkan akun Anda' is displayed. The login form includes two input fields: 'Username' and 'Password' (with a toggle eye icon). A pink 'Login' button with a key icon is at the bottom of the form. The Windows taskbar at the bottom shows the date and time as 22:23 on 28/05/2025.

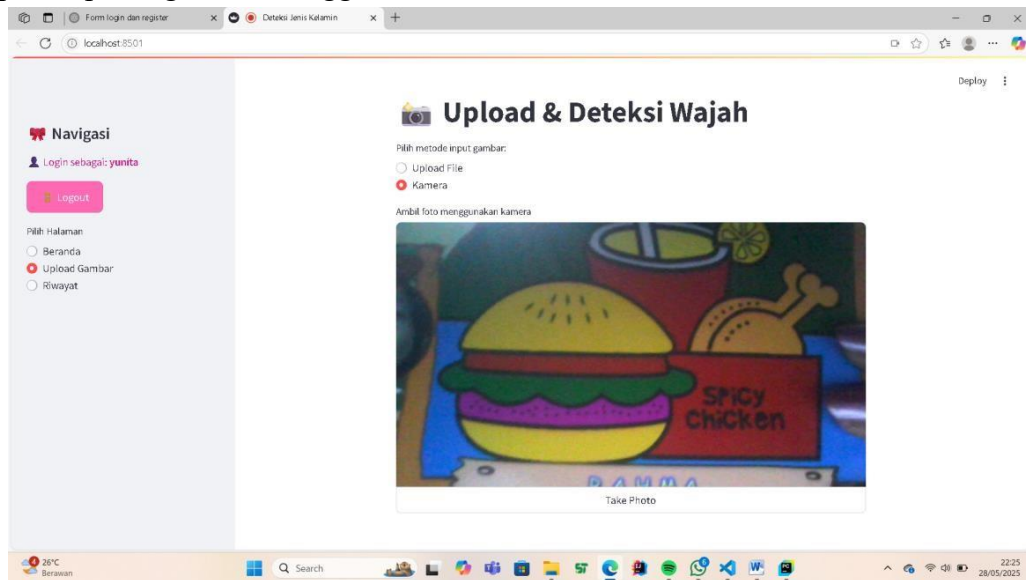
Tampilan beranda,



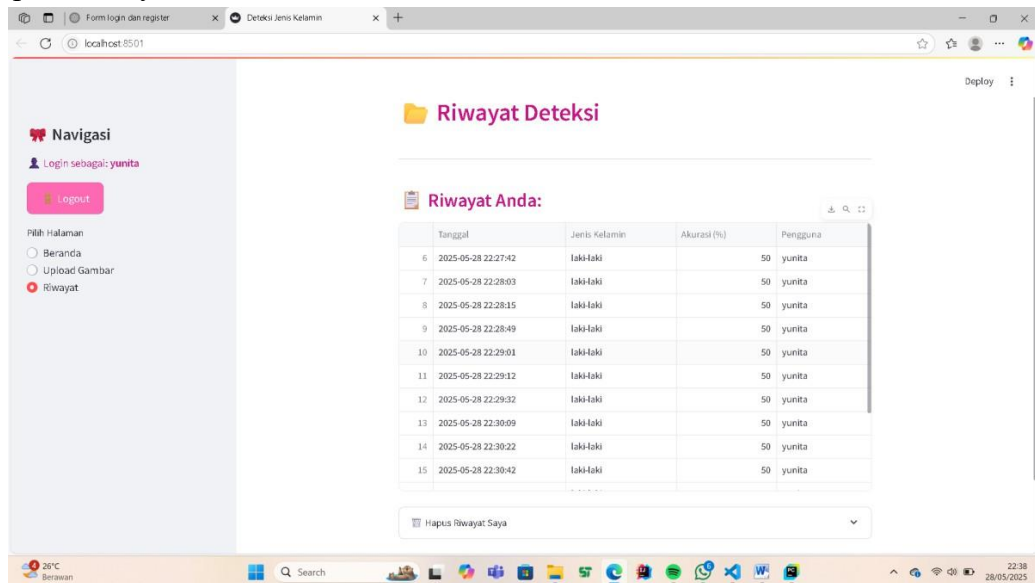
Tampilan upload gambar menggunakan file,

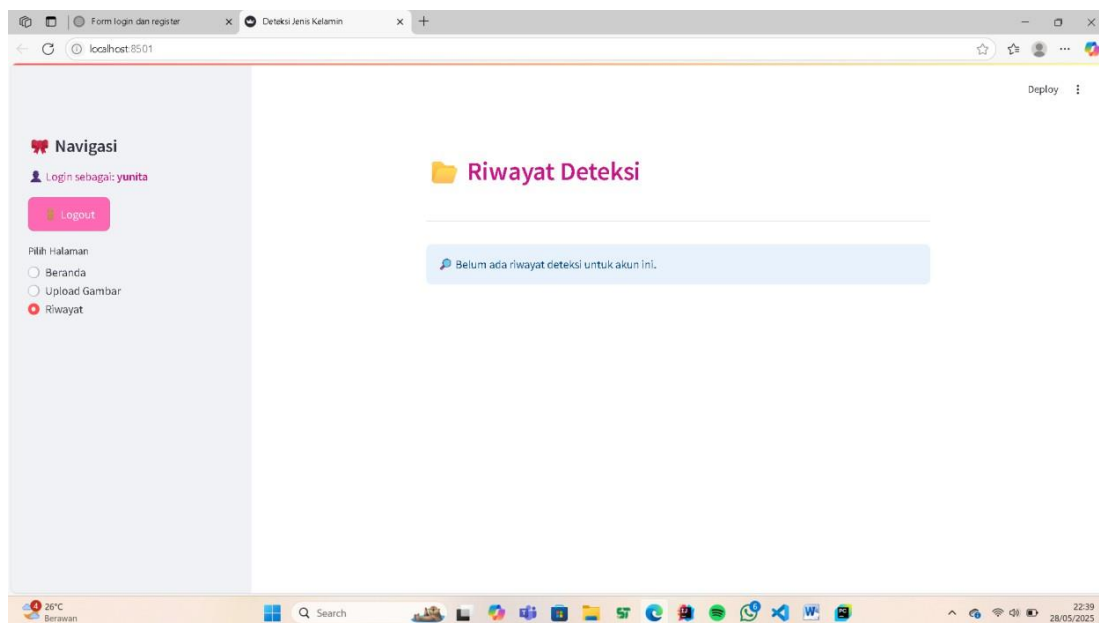
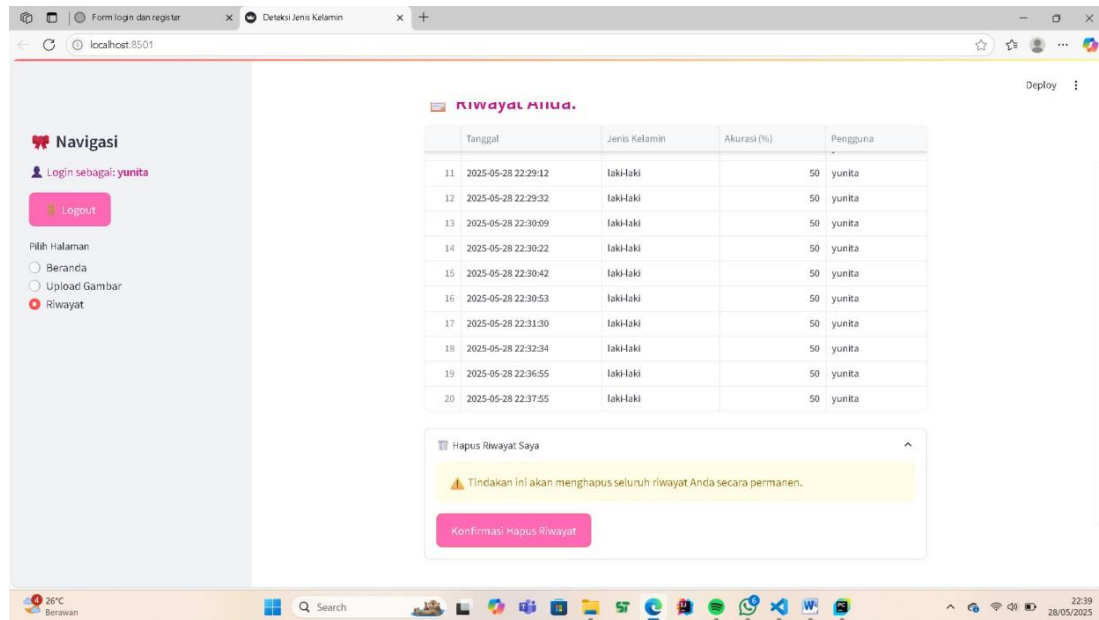


Tampilan upload gambar menggunakan camera,



Tampilan riwayat,





L. Evaluasi Akurasi Model

1. Model yang digunakan: Support Vector Machine (SVM) dengan kernel default
2. Dataset dummy digunakan sebagai contoh (harus diganti dengan dataset nyata seperti UTKFace)
3. Akurasi prediksi pada dataset uji belum tersedia karena model saat ini hanya dummy
4. Untuk penggunaan riil, evaluasi dapat dilakukan dengan:
 - Train-test split (80:20)
 - Cross-validation (misal 5-fold)

- Mengukur metrik seperti accuracy, precision, recall, dan F1-score

Contoh hasil evaluasi :

Wajah berhasil terdeteksi.

Proses Deteksi

Jenis Kelamin: laki-laki

Keakuratan: 50.0%

M. Rekomendasi Pengembangan Selanjutnya

1. Mengganti model dummy dengan model yang dilatih menggunakan dataset nyata
2. Menambahkan facial landmark detector untuk ekstraksi fitur yang lebih akurat
3. Integrasi ke cloud (misal: deployment ke HuggingFace Spaces atau Streamlit Cloud)
4. Menambahkan fitur pelatihan ulang model secara online
5. Menyediakan API REST untuk prediksi melalui sistem eksternal

BAB IV

PENUTUP

Kesimpulan

Berdasarkan penelitian dan pengembangan system pendeteksi jenis kelamin berbasis geometri wajah, dapat disimpulkan bahwa :

1. Geometri wajah merupakan fitur yang cukup representatif untuk membedakan jenis kelamin seseorang secara visual, terutama dengan memanfaatkan titik-titik landmark pada wajah.
2. Dengan menggunakan bahasa pemrograman Python dan pustaka seperti Dlib dan OpenCV, sistem mampu mendeteksi wajah, mengekstraksi fitur geometris, dan mengklasifikasikan jenis kelamin secara otomatis.
3. Algoritma klasifikasi seperti Support Vector Machine (SVM) terbukti efektif untuk melakukan klasifikasi jenis kelamin berdasarkan fitur-fitur yang telah diekstraksi.
4. Penggunaan PyCharm sebagai IDE memberikan kemudahan dalam pengembangan, debugging, dan pengujian sistem secara terstruktur.
5. Sistem yang dibangun menunjukkan performa yang cukup baik pada dataset pengujian, sehingga dapat diimplementasikan pada aplikasi nyata dengan beberapa pengembangan lebih lanjut.

Saran

Untuk pengembangan lebih lanjut, berikut beberapa saran yang dapat dilakukan:

1. Perluasan Dataset: Gunakan dataset yang lebih besar dan beragam dari segi usia, ras, ekspresi, dan pencahayaan untuk meningkatkan generalisasi sistem.
2. Peningkatan Akurasi: Coba terapkan teknik deep learning seperti Convolutional Neural Network (CNN) dengan input dari landmark atau citra mentah sebagai komplementer.
3. Antarmuka Pengguna: Tambahkan antarmuka berbasis GUI (misalnya dengan PyQt atau Tkinter) agar sistem lebih mudah digunakan oleh pengguna non-teknis.
4. Pengujian Nyata: Lakukan uji coba langsung menggunakan kamera real-time untuk menguji performa sistem dalam kondisi dunia nyata.

DAFTAR PUSTAKA

Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2017). Learning Deep Representation for Gender Classification. *IEEE Transactions on Image Processing*, 26(9), 4353–4364.

Dong, X., Yan, S., & Liu, Y. (2019). A gender classification system based on facial geometric features. *Journal of Visual Communication and Image Representation*, 60, 380–389.

Kazemi, V., & Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Sajjad, M., Khan, S., Muhammad, K., & Baik, S. W. (2018). Gender recognition using region specific features and random forest classifier. *Multimedia Tools and Applications*, 77(3), 4025–4042.

King, D. E. (2009). Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research*, 10, 1755–1758.

OpenCV Documentation. (2024). <https://docs.opencv.org>

Dlib C++ Library. (2024). <http://dlib.net>

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.

Turk, M., & Pentland, A. (1991). Face recognition using eigenfaces. *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 586–591.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 511–518.

UTKFace Dataset. (n.d.). Retrieved from <https://susanqq.github.io/UTKFace/>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

OpenCV. (n.d.). Open Source Computer Vision Library. Retrieved from <https://opencv.org/>

Streamlit. (n.d.). Streamlit Documentation. Retrieved from <https://docs.streamlit.io/>

McKinney, W. (2010). Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference, 51–56. (pandas)

NumPy Developers. (2023). NumPy User Guide. Retrieved from <https://numpy.org/doc/>

Python Software Foundation. (n.d.). Python Language Reference. Retrieved from <https://docs.python.org/>