

# Algoritma Machine Learning untuk Pencarian dan Kecocokan Pasangan Berbasis Preferensi Visual dan Obrolan

Designed for: Machine Learning SD-AI    Designed by: Kelompok 2    Date: 20 Maret 2025    Iteration: 1

<div><div>PREDICTION TASK<div>?</div></div><div><p><b>Jenis Tugas:</b></p><ol style="list-style-type: none"><li>Mengelompokkan preferensi visual pengguna berdasarkan pola gambar (clustering)</li><li>Mencari kesamaan gaya komunikasi pengguna (similarity)</li></ol><p><b>Apa yang Diprediksi?</b></p><ol style="list-style-type: none"><li>Kecocokan pasangan berdasarkan preferensi visual</li><li>Cara dan gaya pengguna berkomunikasi dalam chatting</li></ol><p><b>Hasil yang Diprediksi:</b></p><ol style="list-style-type: none"><li>Klaster preferensi visual yang menggambarkan pola ketertarikan pengguna</li><li>Tingkat kesamaan gaya chatting antar pengguna</li></ol><p><b>Kapan Hasilnya Terlihat?</b></p><ol style="list-style-type: none"><li>Setelah pengguna menyelesaikan tes preferensi awal</li><li>Setelah pengguna mulai chatting dengan pasangan yang direkomendasikan</li></ol></div></div>	<div><div>DECISIONS<div>↕</div></div><div><p>Sistem matchmaking ini dirancang untuk merekomendasikan pasangan yang sesuai dengan preferensi pengguna dengan mempertimbangkan berbagai faktor seperti preferensi visual, gaya komunikasi, serta informasi lainnya. Rekomendasi dibuat berdasarkan:</p><ol style="list-style-type: none"><li><b>Menganalisis Preferensi Visual:</b> Menganalisis ketertarikan visual pengguna melalui tes preferensi gambar (wajah, gaya berpakaian, dan suasana dalam foto).</li><li><b>Gaya Komunikasi:</b> Mengevaluasi preferensi chatting pengguna berdasarkan gaya komunikasi, intensitas percakapan, penggunaan emoji, serta kecepatan dalam merespons pesan.</li><li><b>Kriteria Tambahan/ Faktor Non-Visual:</b> Sistem juga mempertimbangkan aspek lain seperti usia, lokasi, status hubungan, agama, preferensi aktivitas harian, nilai-nilai dalam hubungan untuk mendapatkan rekomendasi pasangan yang lebih sesuai.</li></ol></div></div>	<div><div>VALUE PROPOSITION<div>📦</div></div><div><p><b>Cocok untuk:</b> Individu yang sedang mencari pasangan melalui jejaring sosial dengan penyesuaian visual dan obrolan.</p><p><b>Pain Point yang Ingin Ditangani dari Pengguna:</b></p><ol style="list-style-type: none"><li>Ketiadaan algoritma pencarian dan pencocokan berbasis preferensi visual dan obrolan melalui gambar dan <i>chatting</i>.</li><li>Kesulitan individu dalam mengungkapkan preferensi pasangan secara eksplisit dan dalam mencari pasangan idaman yang lebih sesuai</li><li>Masalah ketidakcocokan individu dari segi visual dan/atau obrolan terhadap pasangan hasil <i>match</i>.</li></ol><p><b>Integrasi dengan Workflow:</b></p><ol style="list-style-type: none"><li>Pengguna mengisi preferensi untuk personalisasi (visual, <i>chatting</i>, gaya hidup, dan lain-lain).</li><li>Model mempelajari pola preferensi pengguna dan merekomendasikan pasangan berdasarkan tingkat kecocokan tertinggi.</li><li>Saat interaksi, algoritma mempertimbangkan gaya <i>chatting</i> pengguna untuk memberikan rekomendasi yang lebih personal.</li></ol><p><b>User Interface yang Mendukung:</b></p><ol style="list-style-type: none"><li><i>Dashboard Profile</i> (Rekomendasi Pasangan)</li><li>Match Feed (Pencocokan Berbasis Preferensi).</li><li>Chat Interface (personalisasi gaya komunikasi).</li><li>Feedback Loop (penyesuaian model berdasarkan rating pengguna).</li></ol></div></div>	<div><div>DATA COLLECTION<div>⬇</div></div><div><p>Data awal dalam sistem matchmaking ini diperoleh dari input pengguna saat pendaftaran, termasuk informasi profil, preferensi visual, dan gaya komunikasi. Selain itu, data juga dikumpulkan dari interaksi pengguna seperti pemilihan gambar, pola <i>chatting</i>, serta metadata aktivitas dalam aplikasi. Model machine learning terus diperbarui dengan umpan balik dari interaksi pengguna, di mana algoritma secara otomatis menyesuaikan rekomendasi berdasarkan data terbaru. Untuk menjaga efisiensi biaya dan memastikan data tetap segar, sistem menggunakan pemrosesan batch untuk memperbarui dataset secara berkala, caching untuk mengurangi beban server, serta mekanisme pembaruan manual oleh pengguna. Selain itu, data yang jarang digunakan akan diarsipkan, sementara informasi sensitif dianonimkan guna menjaga privasi pengguna.</p></div></div>	<div><div>DATA SOURCES<div>🗄</div></div><div><p>Data entitas dan hasil yang diamati dalam aplikasi <i>dating</i> ini dapat diperoleh dari sumber internal seperti <b>database pengguna</b> (<i>users</i>) yang menyimpan informasi profil dasar, <b>tabel foto profil</b> (<i>profile_pictures</i>) yang mencatat gambar unggahan beserta fitur visualnya, <b>tabel preferensi visual</b> (<i>visual_preferences</i>) yang merekam interaksi pengguna dengan foto rekomendasi, <b>tabel log obrolan</b> (<i>chat_logs</i>) yang mencatat pola komunikasi pengguna, serta <b>tabel pencocokan</b> (<i>matches</i>) yang menyimpan pasangan yang berhasil dipertemuan dan hasil interaksi mereka. Sementara itu, sumber eksternal dapat mencakup <b>Computer Vision API</b> (Google Vision, OpenCV) untuk mengekstrak fitur visual dari foto profil, <b>Natural Language Processing API</b> (OpenAI, Google NLP) untuk menganalisis gaya <i>chatting</i>, <b>Geolocation API</b> (Google Maps API) untuk mencocokkan pengguna berdasarkan lokasi, serta <b>Social Media API</b> (Instagram, TikTok, Twitter) jika pengguna memberikan izin akses guna memperoleh data preferensi dan interaksi sosial mereka. Dengan menggabungkan data dari sumber internal dan eksternal ini, model machine learning dapat memberikan rekomendasi pasangan yang lebih akurat sesuai dengan preferensi visual dan gaya komunikasi pengguna.</p></div></div>
<div><div>IMPACT SIMULATION<div>✓</div></div><div><p><b>Apa Untung Ruginya Jika Prediksi Benar atau Salah?</b></p><ol style="list-style-type: none"><li><b>Pencocokan Akurat:</b> Pengguna lebih puas, dapat menemukan sesuai preferensi, yang akan membuat pengguna lebih lama dalam menggunakan aplikasi</li><li><b>Pencocokan Tidak Akurat:</b> Pengguna kecewa, tidak menemukan profil yang sesuai, yang akan membuat reputasi aplikasi menurun</li><li><b>Chatting yang Dioptimalkan dengan Baik:</b> Obrolan lebih menarik, pengguna lebih aktif</li><li><b>Chatting yang Dioptimalkan Buruk:</b> Pengguna merasa tidak nyaman atau bosan</li></ol><p><b>Data yang Digunakan untuk Simulasi Sebelum Aplikasi di Deploy:</b></p><p>Dataset dummy tentang preferensi pengguna (gambar dan <i>chatting</i>)</p><p><b>Kapan Sistem Siap di Deploy?</b></p><p>Sistem dianggap siap ketika pencocokan menghasilkan interaksi yang lancar dan pengguna merasa cocok dengan rekomendasi yang diberikan. dimana hasil prediksi dari pelatihan model harus akurat untuk mendapatkan kecocokan yang sesuai.</p></div></div>	<div><div>MAKING PREDICTIONS<div>↔</div></div><div><p><b>Batch atau Real-Time:</b> Prediksi dilakukan secara real-time saat pengguna memperbarui profil atau berinteraksi dalam obrolan. Pembaruan batch juga dilakukan secara berkala.</p><p><b>Frekuensi:</b> Rekomendasi akan terus diperbarui seiring dengan interaksi pengguna (perubahan profi, menyukai profil, aktivitas <i>chatting</i>, mengubah preferensi pencarian).</p><p><b>Waktu Proses:</b> Featurisasi dan pengambilan keputusan berlangsung dalam hitungan detik.</p></div></div>		<div><div>BUILDING MODELS<div>⚙</div></div><div><p><b>1. Model preferensi visual</b> Model menggunakan Unsupervised Learning untuk menemukan rekomendasi pengguna berdasarkan kemiripan visual terhadap beberapa foto pilihan user.</p><p><b>2. Model pencocokan gaya komunikasi</b> Model berbasis NLP dapat menggunakan Transformer untuk mencari similarity teratas. <b>Rekomendasi akhir</b> ditentukan menggunakan Weighted Sum Model.</p><p><b>Durasi komputasi</b> yang dilakukan secara real time saat user melakukan input pada sistem dalam hitungan detik hingga model memberikan output.</p><p><b>Pembaruan model</b> akan dilakukan secara berkala, terutama ketika terdapat pengguna baru, dan terdapat improvement model.</p></div></div>	<div><div>FEATURES<div>III</div></div><div><p><b>Fitur</b> yang digunakan sebagai input model baik model preferensi visual maupun pencocokan gaya komunikasi yaitu berupa vector embedding berdimensi 128 atau 512 dimensi tergantung model yang dikembangkan. Model pencocokan gaya komunikasi juga menggunakan fitur tambahan seperti skor formalitas bahasa, rata-rata panjang chat, frekuensi penggunaan emoji, dan penggunaan bahasa slang.</p><p><b>Transformasi vector embedding</b> data visual diperoleh melalui proses ekstraksi fitur, seperti model deep learning FaceNet atau ArcFace yang dapat mengambil fitur penting dari wajah. Sementara vector embeddings text dapat menggunakan model transformer seperti BERT, RoBERTa, atau Sentence-BERT.</p></div></div>
<div><div><b>Bagaimana Sistem Dijaga agar Adil?</b></div><div><ol style="list-style-type: none"><li>Tidak ada preferensi yang membatasi berdasarkan faktor tertentu.</li><li>Pengguna bisa mengatur sendiri kriteria pasangan yang mereka inginkan.</li></ol></div></div>	<div><div>MONITORING<div></div></div><div><p><b>Metrics &amp; KPIs Untuk Pengguna:</b></p><ol style="list-style-type: none"><li><i>Match success rate</i> (Persentase match yang berlanjut ke interaksi)</li><li><i>Conversation Engagement</i> (Rata-rata jumlah pesan yang dipertukarkan)</li><li><i>User Retention</i> (Persentase pengguna yang tetap aktif setelah pertama kali menggunakan)</li><li><i>User Satisfaction</i> (<i>Feedback</i> terkait kualitas match)</li></ol></div></div>	<div><div><b>Metrics &amp; KPIs Untuk Bisnis:</b></div><div><ol style="list-style-type: none"><li><i>Conversion rate</i> (Pengguna yang berlangganan setelah match)</li><li><i>Churn rate</i> (Persentase pengguna yang meninggalkan platform)</li><li><i>Processing time</i> (Waktu pemrosesan rekomendasi)</li></ol></div></div>	<div><div><b>Frekuensi Review:</b></div><div><ol style="list-style-type: none"><li>Harian (Evaluasi <i>engagement match rate</i>)</li><li>Mingguan (Analisis tren preferensi pengguna)</li><li>Bulanan (Evaluasi <i>user retention</i> dan efektivitas hasil pencocokan)</li></ol></div></div> <div><div></div><div>dan</div></div>	