**PROPOSAL PROJECT PEMBELAJARAN MESIN**

|  |
| --- |
| **A. JUDUL** |
| Penerapan Algoritma K-Nearest Neighbor (KNN) Untuk Klasifikasi Jenis Kelamin Manusia Berdasarkan Objek Wajah |
| **B. RINGKASAN** |
| Jenis kelamin sering menjadi pembeda peran dan tugas dalam kehidupan sehari-hari maupun dalam hal pekerjaan. Sebagian besar model klasifikasi jenis kelamin menggunakan gambar seluruh wajah. Pemrosesan gambar berkaitan dengan pemrosesan fitur yang melibatkan representasi terperinci seperti rambut, dahi, hidung, dan bibir. Tujuan dari penelitian ini adalah untuk mendapatkan tingkat akurasi terbaik dalam proses klasifikasi wajah dalam menentukan jenis kelamin seseorang. Langkah awal penelitian adalah menyiapkan kumpulan data objek wajah yang akan dianalisis. Hasil penelitian yang dilakukan terhadap objek wajah berdasarkan panjang rambut, lebar panjang dahi, lebar panjang hidung, dan ketebalan bibir dengan menggunakan metode K-Nearest Neighbor, dapat disimpulkan bahwa metode ini termasuk dalam algoritma yang sangat baik untuk diterapkan pada klasifikasi objek wajah berdasarkan panjang rambut, lebar panjang dahi, lebar panjang hidung, dan ketebalan bibir secara akurat. Hasil penelitian ini diharapkan dapat menjadi informasi bagi penelitian selanjutnya dalam klasifikasi jenis kelamin manusia menggunakan algoritma K-Nearest Neighbor (KNN). |
| **C. KATA KUNCI** |
| Jenis Kelamin Manusia; K-Nearest Neighbor; Objek Wajah; Klasifikasi; Algoritma |
| **D. PENDAHULUAN** |
| **Latar Belakang**. Wajah atau muka adalah bagian dari kepala, pada manusia meliputi wilayah dari dahi hingga dagu, termasuk rambut, dahi, alis, mata, hidung, pipi, mulut, bibir, gigi, kulit dan dagu(1). Wajah digunakan untuk ekspresi wajah, penampilan serta identitas(2). Pengenalan pola wajah dapat dipandang untuk mengenali emosi, ras, ataupun pemilik wajah tersebut berdasarkan fitur-fitur wajah yang dimiliki(3).  Klasifikasi paling mendasar mengenai manusia adalah jenis kelamin dari manusia tersebut. Gender adalah klasifikasi gramatikal terhadap kata- kata dan kata-kata terkait lainnya yang berkaitan dengannya yang secara garis besar berhubungan dengan keberadaan dua jenis kelamin atau kenetralan(4). Gender juga merupakan suatu konstruksi budaya yang sifatnya terbuka terhadap bagi segala perubahan(5). Sebagian besar model klasifikasi jenis kelamin menggunakan gambar seluruh wajah. Penelitian ini melakukan pemrosesan fitur yang melibatkan representasi terperinci seperti rambut, dahi, hidung, dan bibir(6). Pendeteksian objek wajah merupakan salah satu tahap awal yang sangat penting sebelum dilakukan proses pengenalan wajah. Pada saat ini kebutuhan untuk memperkirakan jenis kelamin seseorang dibutuhkan dengan beberapa cara otomatis bermunculan. Namun hal ini masih sangat sulit dilakukan oleh komputer karena harus memperkirakan jenis kelamin manusia beardasarkan gambar wajah.  Algoritma K-Nearest Neighbor (KNN) akan mengklasifikasikan objek wajah uji kedalam kelas dengan jumlah anggota terbanyak. Tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan sample dari data lama(7). Atribut yang sudah dipilih nantinya akan sebagai parameter untuk menentukan kelas(8). KNN mengikuti strategi “*bird of a feather*” dalam menentukan dimana data baru sebaiknya ditempatkan(9). Algoritma KNN mengasumsikan bahwa sesuatu yang mirip akan ada dalam jarak yang berdekatan atau bertetangga(10). Hal tersebut berarti data-data yang cenderung serupa akan dekat satu sama lain.  Tujuan dalam penelitian deteksi wajah ini adalah untuk meningkatkan akurasi deteksi agar dapat mengenali wajah dengan lebih tepat, khususnya membedakan jenis kelamin dari wajah seseorang. Hal ini melibatkan pengembangan dan evaluasi algoritma atau model. Dari proses tersebut diharapkan mampu mengenali wajah dengan lebih baik dalam berbagai kondisi.  **Rumusan Permasalahan.** Penelitian mengenai klasifikasi jenis kelamin seseorang berdasarkan objek wajah seperti panjang rambut, lebar panjang dahi, lebar panjang hidung, dan ketebalan bibir belum pernah ada sebelumnya. Oleh karena itu, saya selaku peneliti berusaha untuk membuat klasifikasi jenis kelamin seseorang dengan menggunakan objek yang baru. Sistem pengenalan wajah membutuhkan sebuah fitur dari suatu objek untuk dapat dikenali, kemudian fitur tersebut akan dicocokkan dengan fitur objek lainnya.  **Pendekatan Pemecahan Masalah.** Dalam konteks ini, penelitian bertujuan menerapkan algoritma KNN untuk mengetahui bagaimana tingkat akurasi klasifikasi jenis kelamin. Dengan demikian, penelitian ini diharapkan guna membangun sistem yang bisa mengklasifikasi jenis kelamin diantara laki-laki dan perempuan hanya dengan menggunakan objek wajah. Algoritma KNN digunakan untuk mengklasifikasikan objek baru berdasarkan atribut dari data lama. Atribut yang sudah dipilih tersebut nantinya akan digunakan sebagai pijakan untuk menentukan data baru sebaiknya ditempatkan. Setelah itu peneliti dapat menyimpulkan hasil klasifikasi jenis kelaminnya.  **State of the Art.** Sejumlah studi pendahuluan telah dilakukan untuk mengidentifikasi penggunaan algoritma KNN dalam klasifikasi jenis kelamin seseorang antara lain :   1. “Implementasi K-Nearest Neighbor Untuk Klasifikasi Jenis Kelamin Berdasarkan   Analisis Citra Wajah”  Penelitian ini melakukan klasifikasi jenis kelamin berdasarkan citra wajah manusia. Teknik olah citra tersebut meliputi modifikasi kecemerlangan, negasi, peningkatan kontras, dan thresholding. Pengenalan citra yang dilakukan pada penelitian ini yaitu melakukan metode morfologi untuk proses segmentasi ekstraksi citra dan metode klasifikasi K-Nearest Neighbor untuk memperkuat tingkat keakuratan data hasil pengolahan citra.   1. “Perbandingan Metode K-Nearest Neighbor dan Naïve Baiyes Untuk Klasifikasi   Gender Berdasarkan Mata”  Penelitian ini melakukan perbandingan algoritma K-Nearest Neighbor dan Naïve Baiyes untuk klasifikasi gender berdasarkan mata. *Histogram of Oriented Gradient* (HOG) merupakan metode yang digunakan untuk Image Processing pada penlitian ini yang bertujuan untuk mendeteksi objek. KNN digunakan untuk metode klasifikasi yang menentukan kategori berdasarkan mayoritas kategori. Sedangkan Naïve Baiyes Classifier (NBC) adalah metode untuk menghitung sekumpulan probabilitas dengan menambahkan kombinasi frekuensi dan nilai dalam kumpulan data yang diberikan.  Kedua penelitian diatas sama sama menggunakan algoritma K-Nearest Neighbor (KNN) untuk klasifikasi jenis kelamin. Dengan menggabungkan kedua penelitian tersebut peneliti bisa mendapatkan gambaran dan melakukan penelitian yang lebih akurat.  **Kebaruan.** Kebaruan dalam penelitian ini terletak pada objek yang akan dianalisis yaitu objek wajah berdasarkan panjang rambut, lebar panjang dahi, lebar panjang hidung, dan ketebalan bibir berbasis algoritma K-Nearest Neighbor.  **Roadmap Penelitian.** Penelitian ini dibuat dengan runtunan sebagai berikut :   1. Menyiapkan dataset gambar objek wajah yang akan dianalisa. Dataset meliputi, apa-   kah rambut panjang atau tidak, berapa lebar dahi dari kanan ke kiri dalam satuan cm, berapa lebar dahi dari tempat tumbuhnya rambut sampai ke alis dalam satuan cm, apakah hidung lebar atau tidak, apakah hidung panjang atau tidak, apakah bibir tipis atau tidak, dan terakhir apakah individu tersebut memiliki jarak yang panjang dari hidung ke bibir atau tidak.   1. Setelah data disiapkan, langkah selanjutnya adalah memasukkan data gambar ke dal-   am 1 folder agar data gambar bisa dicari dan diolah dengan mudah.   1. Dari setiap gambar tersebut akan dilakukan segmentasi untuk memisahkan hasil objek wajah seperti berapa lebar dahi dari kanan ke kiri dan seterusnya. 2. Setelah data objek wajah didapatkan dari hasil segmentasi, langkah selanjutnya adalah penerapan klasifikasi dengan metode K-Nearest Neighbor dari data yang didapat. 3. Selesai melakukan proses klasifikasi menggunakan metode K-Nearest Neighbor, maka target atau hasilnya akan terlihat yaitu antara laki laki ataupun perempuan. |
| **E. METODE** |
| **1. Studi Pustaka**  Studi Pustaka dilakukan dengan tahapan pengerjaan model machine learning. Tahapan pengerjaan ini akan dideskripsikan melalui diagram alir. Pengerjaan model machine learning penulisan codenya menggunakan bahasa pemrograman *Python*, sedangkan untuk tahapan *preprocessing* hingga pemodelan menggunakan software Visual Studio Code.  Gambar 1. Tahapan Pengerjaan Model  **2.** **Preprocessing**  Tahapan preprocessing data digunakan untuk menghilangkan beberapa permasalahan yang ada pada data guna meningkatkan kualitas data dan menghilangkan beberapa permasalahan yang bisa mengganggu saat pemrosesan data.  **3. Dataset yang digunakan**   1. Gender Classification Dataset (Dataset Publik)   Dataset didapatkan dari Kaggle (<https://www.kaggle.com/datasets/elakiricoder/gender-classification-dataset>, diakses 26 April 2024) data ini sangat cocok digunakan untuk pemula yang ingin mencoba memecahkan masalah klasifikasi dalam *machine learning.* Data terdiri dari 7 fitur dan 5002 *record*. Berikut atribut dari dataset yang akan digunakan.  Tabel 1. Deskripsi Dataset   |  |  | | --- | --- | | Nama Atribut | Deskripsi | | long\_hair | value atribut hanya 0 dan 1, dimana 1 untuk mewakili “rambut yang panjang” dan 0 untuk “rambut yang tidak panjang” | | forehead\_width\_cm | lebar dahi dalam satuan centimeter (cm) | | forehead\_height\_cm | tinggi dahi dalam satuan centimeter (cm) | | nose\_wide | value atribut hanya 0 dan 1, dimana 1 untuk mewakili “hidung lebar” dan 0 untuk “hidung tidak lebar” | | nose\_long | value atribut hanya 0 dan 1, dimana 1 untuk mewakili “hidung panjang” dan 0 untuk “hidung tidak panjang” | | lips\_thin | value atribut hanya 0 dan 1, dimana 1 untuk mewakili “bibir yang tipis” dan 0 untuk “bibir tidak tipis” | | distance\_nose\_to\_lip\_long | value atribut hanya 0 dan 1, dimana 1 untuk mewakili “jarak yang jauh antara hidung dan bibir” dan 0 untuk “jarak antara hidung dan bibir tidak jauh” | | gender | output atau hasil dengan value “Male” atau “Female” |   **4. Normalisasi**  Tahapan ini akan mengganti nilai dari atribut gender guna mempermudah pemrosesan data. Nilai dari atribut yang sebelumnya “male” dan “female” akan diganti menjadi numerik, yaitu 1 untuk mewakili “male” dan 0 untuk mewakili “female”.  **5. Pengukuran Jarak Euclidean**  Setelah mendapatkan data wajah, langkah berikutnya adalah melakukan perhitungan jarak Euclidean dari dataset. Adapun rumus menghitung Euclidean adalah :  keterangan :  xi = Data Pengujian  yi = Data Pelatihan  i = Variabel Data  d = Jarak  Lalu jika sudah mendapatkan hasil jarak Euclidean, langkah selanjutnya adalah mengurutkan hasil perhitungan jarak Euclidean dari jarak yang paling kecil. Jika sudah, langkah berikutnya adalah mengambil data sebanyak nilai K. Dalam algoritma KNN, pendekatan sederhana dalam menentukan nilai k bisa dilakukan dengan rumus :  K =  dimana n merupakan banyaknya data yang dihitung.  Langkah selanjutnya adalah mengumpulkan kategori y, dimana y merupakan data jarak euclidean terkecil yang paling banyak terkumpul. Lalu penentuan hasil klasifikasi dari beberapa kategori y yang terkumpul, dimana data yang paling mayoritas merupakan hasil klasifikasi yang di dapat.  **6. Eksperimen**  Dalam melakukan eksperimen, pengujian ini menggunakan device laptop VivoBook Asus X515JAB dengan spesifikasi RAM 4GB, SSD NVME 512GB dan Processor Intel Core i3 Gen 10.  **6. Luaran Penelitian**  Berikut luaran penelitian ini selama 14 minggu (Tabel 1):  Tabel 2. Luaran Penelitian   |  |  |  | | --- | --- | --- | |  | Luaran Wajib | Luaran Tambahan | | 14 minggu | Model Machine Learning (python ~ .ipynb) | - | |
| **F. JADWAL PROJECT** |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | No | Nama Kegiatan | Minggu ke- | | | | | | | | | | | | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | | | 1 | Mempelajari dan mencari algoritma yang cocok |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | | 2 | Membuat dan submit Proposal |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | | 3 | Pencarian dataset yang simpel |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | | 4 | Mempelajari Jupiter Notebook dan mencari referensi jurnal |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | | 5 | Eksperimen dan membuat model machine learning |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | |
| **G. DAFTAR PUSTAKA** |
| 1. Hardiyanto D, Anggun Sartika D. Optimalisasi Metode Deteksi Wajah berbasis Pengolahan Citra untuk Aplikasi Identifikasi Wajah pada Presensi Digital. Setrum Sist Kendali-Tenaga-elektronika-telekomunikasi-komputer [Internet]. 2018 Jun 11;7(1):107. Available from: https://jurnal.untirta.ac.id/index.php/jis/article/view/3367  2. Budi A, Suma’inna S, Maulana H. Pengenalan Citra Wajah Sebagai Identifier Menggunakan Metode Principal Component Analysis (PCA). J Tek Inform [Internet]. 2018 Jan 15;9(2). Available from: http://journal.uinjkt.ac.id/index.php/ti/article/view/5608  3. Widiakumara IKS, Putra IKGD, Wibawa KS. Aplikasi Identifikasi Wajah Berbasis Android. Lontar Komput J Ilm Teknol Inf. 2017;8(3):200.  4. Fakih M. Analisis Gender dan Transformasi Sosial, Yogyakarta. Insist Press; 1996.  5. Kurniawan C, Irsyad H. Perbandingan Metode K-Nearest Neighbor Dan Naïve Bayes Untuk Klasifikasi Gender Berdasarkan Mata. J Algoritm. 2022 Apr 11;2(2):82–91.  6. Triwijoyo BK. Model Fast Tansfer Learning pada Jaringan Syaraf Tiruan Konvolusional untuk Klasifikasi Gender Berdasarkan Citra Wajah. MATRIK J Manajemen, Tek Inform dan Rekayasa Komput [Internet]. 2019 May 29;18(2):211–21. Available from: https://journal.universitasbumigora.ac.id/index.php/matrik/article/view/376  7. Muchtar H, Said F. Sistem Identifikasi Plat Nomor Kendaraan Menggunakan Metode Robert Filter dan Framing Image Berbasis Pengolahan Citra Digital. Resist (elektRonika kEndali Telekomun tenaga List kOmputeR) [Internet]. 2019 Nov 3;2(2):105. Available from: https://jurnal.umj.ac.id/index.php/resistor/article/view/4792  8. Prahudaya TY, Harjoko A. METODE KLASIFIKASI MUTU JAMBU BIJI MENGGUNAKAN KNN BERDASARKAN FITUR WARNA DAN TEKSTUR. J Teknosains [Internet]. 2017 Aug 30;6(2):113. Available from: https://jurnal.ugm.ac.id/teknosains/article/view/26972  9. Fansyuri M, Yunita D. KLIK: Kajian Ilmiah Informatika dan Komputer Implementasi K-Nearest Neighbor Untuk Klasifikasi Jenis Kelamin Berdasarkan Analisis Citra Wajah. Media Online). 2023;3(6):1208–16.  10. Devita RN, Herwanto HW, Wibawa AP. Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia. J Teknol Inf dan Ilmu Komput. 2018;5(4):427–34. |