

KLASIFIKASI GENDER MENGGUNAKAN DATA WAJAH DENGAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBORS

Adittia Fathah^{*1}, Christina Juliane²

^{1,2}Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) LIKMI Bandung, Bandung
Email: ¹ adittiafathah11@outlook.co.id, ² christina.juliane@likmi.ac.id

*Penulis Korespondensi

(Naskah masuk: 12 Februari 2024, diterima untuk diterbitkan: 10 Februari 2025)

Abstrak

Identifikasi *gender* saat ini lebih sulit dilakukan. Penyebabnya antara lain banyaknya klasifikasi *gender*, penggunaan identitas palsu di media sosial dan semakin maraknya foto palsu. Peristiwa nyata yang terjadi adalah banyaknya klasifikasi di negara Thailand yang memiliki 18 *gender*. Peristiwa lainnya adalah penambahan *gender* "X" pada aplikasi permohonan *passport* di Amerika dan beredarnya foto palsu yang diedit dengan aplikasi *FaceApp*. Kejadian tersebut menyebabkan perlunya membuat model yang bisa melakukan klasifikasi *gender* agar *gender* asli dari seseorang bisa diketahui. Penelitian dilakukan dengan mencari model yang bisa mengklasifikasikan *gender*. Caranya adalah dengan membandingkan hasil akurasi dua algoritma yaitu *Naïve Bayes* dan K-NN (*K-Nearest Neighbors*). Metode yang digunakan mengikuti tahapan dalam KDD (*Knowledge Discovery in Database*). Atribut yang dipakai adalah bagian-bagian pada wajah yaitu lebar dahi, lebar hidung, panjang hidung, bibir dan jarak hidung ke bibir. Akurasi kedua algoritma diuji dengan metode *Cross Validation* dan *Confusion Matrix*. Tujuan penelitian ini adalah membuat model klasifikasi *gender* dengan menggunakan data wajah serta menentukan model yang lebih baik antara *Naïve Bayes* atau K-NN. Hasil pengujian menunjukkan, kedua algoritma memiliki akurasi yang sangat baik. Namun algoritma *Naïve Bayes* memiliki nilai AUC yang lebih tinggi yaitu 0,996 dibanding algoritma K-NN yang memiliki nilai AUC sebesar 0,992. Berdasarkan nilai tersebut, atribut bagian-bagian pada wajah yaitu lebar dahi, lebar hidung, panjang hidung, bibir dan jarak hidung ke bibir dapat digunakan untuk klasifikasi *gender*, karena bisa menghasilkan akurasi yang baik. Namun, model *Naïve Bayes* lebih direkomendasikan karena nilai akurasinya lebih tinggi dan stabil. Selain itu, model yang dihasilkan juga lebih akurat, karena data yang digunakan adalah hasil pengukuran dan bentuk wajah secara fisik, bukan berupa foto wajah yang dapat dimodifikasi.

Kata kunci: *Data Mining, Klasifikasi, Gender, Wajah, Naïve Bayes, K-NN, Cross Validation, Confusion Matrix.*

GENDER CLASSIFICATION USING FACE DATA WITH THE NAÏVE BAYES ALGORITHM AND K-NEAREST NEIGHBORS ALGORITHM

Abstract

Gender identification is more difficult today. The reasons include the many gender classifications, the use of fake identities on social media and the proliferation of fake photos. An obvious case in point is the many classifications in the country of Thailand, which has 18 genders. Other incidents include the addition of gender "X" to passport applications in America and the circulation of fake photos edited with the FaceApp app. These events led to the need to create a model that can perform gender classification so that the true gender of a person can be known. The research was conducted by finding a model that can classify gender. The trick is to compare the accuracy results of two algorithms, namely Naïve Bayes and K-NN (K Nearest Neighbor). The method used follows the stages in KDD (Knowledge Discovery in Database). The attributes used are parts of the face, namely forehead width, nose width, nose length, lips and nose-to-lip distance. The accuracy of both algorithms was tested using Cross Validation and Confusion Matrix methods. The purpose of this research is to create a gender classification model using facial data and determine the better model between Naïve Bayes or K-NN. The test results show that both algorithms have very good accuracy. However, the Naïve Bayes algorithm has a higher AUC value of 0.996 than the K-NN algorithm which has an AUC value of 0.992. Based on these values, the attributes of the parts of the face, namely forehead width, nose width, nose length, lips and nose to lip distance can be used for gender classification, because they can produce good accuracy. However, the Naïve Bayes model is more recommended because the accuracy value is higher and more stable. In addition, the resulting model is also more accurate, as the data used is the result of measurements and the physical shape of the face, rather than a photo of the face that can be modified.

Keywords: *Data Mining, Classification, Gender, Face, Naïve Bayes, K-NN, Cross Validation, Confusion Matrix*

1. PENDAHULUAN

Identifikasi *gender* saat ini lebih sulit dilakukan. Penyebabnya antara lain banyaknya klasifikasi *gender*, penggunaan identitas palsu di media sosial dan semakin maraknya foto palsu. Dilansir dari situs berita *Sindo News*, negara dengan *gender* terbanyak adalah Thailand dengan jumlah *gender* sebanyak 18. Negara Thailand memiliki beragam identitas *gender*, termasuk laki-laki (laki-laki yang tertarik pada perempuan), perempuan (perempuan yang tertarik pada laki-laki), *tom* (perempuan yang berpakaian seperti laki-laki dan tertarik pada perempuan *dee*), *dee* (perempuan yang tertarik pada perempuan *tom* atau *dee*), *tom gay* (perempuan yang tertarik pada perempuan, *tom*, dan *dee*), *raja tom gay* (*tom* maskulin yang tertarik pada *tom*), *bi* (perempuan yang tertarik pada biseksual, *tom*, lesbian, dan laki-laki), *kapal* (laki-laki yang tertarik pada perempuan, *gay kings*, dan *gay queens*), *ratu gay* (laki-laki feminis yang tertarik pada laki-laki), *raja gay* (laki-laki maskulin yang tertarik pada laki-laki), *tom gay* dua arah (*tom* yang bisa menjadi *tom gay king* atau *tom gay queen*), *ratu tom gay* (*tom* feminin yang tertarik pada *tom*), lesbian (perempuan yang tertarik pada perempuan), *kathoe* atau *ladyboy* (laki-laki yang ingin menjadi perempuan), *adam* (laki-laki yang tertarik pada *tom*), *angee* (*kathoe* yang tertarik pada *tom*), *ceri* (perempuan yang tertarik pada *gay* dan *kathoe*), dan *samyaaan* (perempuan yang tertarik pada *tom*, lesbian, perempuan, dan bisa menjadi salah satu dari mereka (Fauziah, 2023). Contoh lain terjadi di Amerika Serikat (AS) ketika Menteri Luar Negerinya Antony Blinken menyampaikan kebijakan bahwa warga negaranya dapat memilih pilihan ketiga selain laki-laki dan perempuan yaitu “X” pada aplikasi paspor AS yang menandakan bukan laki-laki atau perempuan (DW, 2022). Dampak yang ditimbulkan akibat banyaknya klasifikasi gender adalah sulit menentukan apakah orang tersebut benar-benar laki-laki atau perempuan.

Media sosial dalam kehidupan sehari-hari sudah menjadi bagian penting. Media sosial biasanya digunakan melakukan interaksi dengan orang lain, namun ada juga yang menggunakannya untuk membuat akun palsu dan melakukan tindakan ilegal. Media sosial juga memberi kemudahan untuk menyembunyikan atau memberi informasi yang salah (Wani et al., 2019). Kedua hal tersebut menyebabkan identitasnya sulit diidentifikasi termasuk jenis kelamin aslinya. Kejahatan yang dapat dilakukan dengan akun palsu seperti pelecehan dunia maya, perundungan, penguntitan, peperangan dan terorisme (Hussein et al., 2019).

Foto juga turut menambah sulitnya identifikasi *gender*. Saat ini, semakin banyak aplikasi yang menyediakan fitur edit foto secara instan. Salah satu contohnya adalah aplikasi *FaceApp*. Aplikasi ini *booming* pada tahun 2017.



Gambar 1 Logo dan Fitur FaceApp
Sumber : (Danar, 2019)

FaceApp memiliki kemampuan untuk mengubah foto wajah menjadi tampak lebih muda atau lebih tua. Bahkan lebih canggihnya lagi, *FaceApp* mampu merubah foto wajah dari laki-laki menjadi seperti perempuan maupun sebaliknya dalam satu klik saja. Tercatat aplikasi ini sudah di unduh lebih dari 100 juta kali di *Google Play Store*.

Gender adalah salah satu atribut sosial yang penting dalam kehidupan manusia. Identifikasi *gender* dapat memiliki implikasi yang signifikan dalam berbagai aspek, termasuk budaya, sosial, psikologis, dan ekonomi. Namun, dalam beberapa situasi identifikasi *gender* mungkin sulit atau tidak jelas, seperti dalam konteks data digital atau data yang *anonym*. *Gender* merupakan salah satu informasi yang banyak dibutuhkan dalam melakukan pengenalan seseorang pada berbagai bidang (Devito et al., 2019). Oleh karena itu, telah banyak peneliti yang telah mengembangkan sebuah sistem klasifikasi untuk dapat mengetahui *gender* dari seseorang. Salah satu cara untuk dapat mengenali *gender* dari seseorang adalah dengan melihat wajah dari orang tersebut (Armandhani et al., 2019).

Gender memiliki peran penting dalam interaksi di masyarakat. Bahkan di zaman modern seperti saat ini, penggunaan komputer terus berkembang pesat, sehingga permintaan akan model klasifikasi *gender* semakin meningkat. Banyak aplikasi yang memerlukan sistem identifikasi *gender*, seperti dalam antarmuka interaksi manusia-komputer yang menyesuaikan perilaku perangkat lunak dengan jenis kelamin pengguna, pengumpulan data demografis untuk menganalisis tren dan merekomendasikan produk berdasarkan jenis kelamin, analisis media sosial (Hussein et al., 2019), (Wani et al., 2019), penelitian pasar dan penentuan segmentasi pelanggan (Goenawan et al., 2019). Klasifikasi *gender* juga dapat dilihat dari gaya berjalan, bentuk iris mata, bentuk rambut dan tangan. Namun, metode yang menonjol adalah didasarkan pada fitur (Wahyu et al., 2019). Klasifikasi *gender* menggunakan teknik dan metode *data mining* untuk mengidentifikasi dan memprediksi jenis kelamin individu berdasarkan atribut atau fitur yang ada. Mudah bagi manusia untuk mengenali *gender* seseorang dengan melihat wajahnya, akan tetapi bagi komputer tidak mudah untuk dapat mengenali *gender* seseorang tanpa

adanya informasi tambahan yaitu ciri-ciri fisik dari orang tersebut (Prayogo, 2022).

Data mining, secara sederhana merupakan proses ekstraksi informasi yang signifikan namun tersirat dan belum terungkap secara eksplisit. Selain itu, *data mining* memiliki kaitan dengan beragam disiplin ilmu seperti statistik, pembelajaran mesin, pengenalan pola, algoritma komputasi, teknologi basis data, dan komputasi berkinerja tinggi.

Klasifikasi merupakan salah satu proses dalam data mining yang bertujuan untuk menemukan pola dan menetapkan atribut atau label kelas dari sampel yang akan dikelompokkan. Tujuan klasifikasi adalah untuk membedakan kelas-kelas data dengan mencari model atau fitur yang dapat memprediksi kelas suatu objek dengan akurat (Ajjah & Kurniawan, 2023). Klasifikasi adalah proses yang penting dalam mengkategorikan objek sesuai dengan kelas-kelas yang telah ditentukan. Salah satu fungsi utama dari klasifikasi adalah memprediksi kelas atau grup tertentu, yang melibatkan pembangunan model berdasarkan data latih dan nilai-nilai label kelas untuk mengklasifikasikan data baru (Febrinamas et al., 2023).

Istilah "*gender*" sering diterjemahkan dari Bahasa Inggris sebagai "jenis kelamin" (*sex*), sehingga sering kali dianggap sebagai sinonim dari jenis kelamin. Namun, dalam berbagai konteks budaya dan sosial, konsep *gender* mengalami perluasan makna yang melibatkan pengaruhnya dalam ranah politik, hukum, ekonomi, dan lainnya. *Gender* merujuk pada karakteristik yang digunakan untuk menggambarkan perbedaan antara laki-laki dan perempuan, yang kemudian mempengaruhi penentuan kedudukan sosial di antara keduanya. Perbedaan jenis kelamin, yaitu laki-laki dan perempuan, tidak hanya terkait dengan aspek biologis, tetapi juga dengan ciri-ciri fisik permanen yang dimiliki masing-masing. Misalnya, perempuan memiliki kemampuan untuk melahirkan, menyusui, serta mengalami menstruasi, sedangkan laki-laki memiliki ciri seperti janggut atau kumis (Azizah, 2021).

Metode *Naive Bayes* merupakan salah satu teknik dalam bidang *data mining* yang menggunakan prinsip probabilitas (kemungkinan) dalam menyelesaikan masalah (Buulolo, 2020). Teknik ini didasarkan pada konsep klasifikasi yang menggunakan perhitungan probabilitas, yang pertama kali diperkenalkan seorang ilmuwan Inggris, Thomas Bayes, untuk memprediksi peluang kejadian berdasarkan pengalaman atau situasi yang telah terjadi sebelumnya. Keunggulan utama dari penggunaan metode *Naive Bayes* adalah kemampuannya untuk memberikan estimasi parameter yang diperlukan dalam proses klasifikasi hanya dengan menggunakan jumlah data latih yang relatif kecil (Prayogo, 2022).

Klasifikasi *Naive Bayes* adalah metode klasifikasi yang menggunakan teorema probabilitas

Bayes untuk mengelompokkan data ke dalam kelas-kelas tertentu. Keunggulan metode ini terletak pada sifat "naif" atau sederhananya, yaitu mengasumsikan bahwa semua fitur yang digunakan untuk klasifikasi bersifat independen satu sama lain, meskipun mungkin ada ketergantungan antar fitur. Walaupun asumsi ini tampak sederhana, kekuatan utama algoritma *Naive Bayes* adalah kemampuannya dalam menangani *dataset* berdimensi tinggi, efisiensinya dalam pemrosesan data, dan ketangguhannya dalam menghadapi ketidakpastian atau variasi dalam data (Fansyuri & Yunita, 2024).

Nilai probabilitas untuk setiap kelas yang dituju akan dihitung dengan memperhitungkan input yang diberikan pada tahap klasifikasi. Kelas target dengan probabilitas tertinggi akan ditetapkan sebagai kelas data input tersebut (Aprilia, 2024).

Metode K-NN merupakan salah satu algoritma dalam bidang *data mining* yang digunakan untuk mengelompokkan objek baru berdasarkan mayoritas kategori tetangga terdekatnya (Buulolo, 2020). Pendekatan K-NN ini memanfaatkan data pembelajaran untuk mengklasifikasikan objek berdasarkan jarak terdekat dengan objek tersebut. Saat diberikan sebuah titik *query*, algoritma K-NN akan mencari sejumlah k objek atau titik pembelajaran yang memiliki jarak paling dekat dengan titik *query* tersebut. Selanjutnya, nilai prediksi untuk *query* akan ditentukan berdasarkan mayoritas klasifikasi tetangganya (Muhadi & Octaviano, 2023). Proses ini melibatkan pembagian data menjadi data latih dan data uji untuk menemukan tetangga terdekat. Setelah pembagian tersebut, jarak antara setiap data uji dengan data latih dihitung (Bahtiar, 2023). Kelebihan utama dari metode K-NN adalah kemampuannya untuk memberikan hasil yang cepat dalam pelatihan, pendekatan yang sederhana, mudah dipahami, dan efektif saat diterapkan pada data latih yang besar (Muhadi & Octaviano, 2023).

Algoritma K-NN adalah metode yang sederhana namun efektif untuk klasifikasi dan regresi, terutama ketika tidak ada asumsi khusus mengenai distribusi data. K-NN bekerja berdasarkan prinsip bahwa data dengan karakteristik serupa cenderung berada dalam kelompok yang sama atau memiliki nilai yang berdekatan. Algoritma ini mengukur jarak antar titik data untuk menentukan seberapa mirip data baru dengan data historis, dan berdasarkan jarak tersebut, algoritma memperkirakan nilai yang diinginkan (Utami & Ayumi, 2024).

Namun, kelemahan utama dari K-NN adalah ketergantungannya pada jumlah data dan kinerjanya yang menurun saat menangani data dengan dimensi tinggi. Selain itu, pemilihan jumlah tetangga (K) memiliki pengaruh signifikan terhadap hasil prediksi dan memerlukan proses penyesuaian (*tuning*) yang cermat (Utami & Ayumi, 2024).

Langkah-langkah untuk mengklasifikasikan data menggunakan algoritma K-NN adalah sebagai berikut (Devian et al., 2024):

- Menentukan nilai K yang akan digunakan.
- Menghitung jarak antara data latih dan data uji.
- Mengelompokkan data berdasarkan hasil perhitungan jarak.
- Mengelompokkan data berdasarkan nilai tetangga terdekat.
- Memilih nilai tetangga terdekat yang paling sering muncul sebagai prediksi untuk data selanjutnya.

Penelitian ini dilakukan untuk mencari model yang bisa digunakan untuk klasifikasi *gender* berdasarkan data wajah dengan cara membandingkan akurasi dari dua algoritma yaitu *Naïve Bayes* dan *K-NN*. Meskipun penelitian yang sama sudah pernah dilakukan oleh (Prayogo, 2022), ada perbedaan dari algoritma yang digunakan. Pada penelitian yang dilakukan sebelumnya, hanya menggunakan satu algoritma yaitu *Naïve Bayes*. Namun pada penelitian ini algoritmanya ditambahkan dengan *K-NN* sebagai pembandingan. Algoritma *Naïve Bayes* dipilih karena atribut yang digunakan tidak saling bergantung satu sama lain (*independen*) dan mampu menghasilkan akurasi sebesar 0,85 berdasarkan penelitian yang dilakukan oleh (Prayogo, 2022), sedangkan algoritma *K-NN* dipilih karena nilai dari setiap atribut bisa dikategorikan (*categorical*) dan memiliki cara yang kerja yang lebih kompleks dari *Naïve Bayes*. Meskipun sumber *dataset* yang digunakan sama dengan penelitian (Prayogo, 2022) yaitu (Issaden, 2020), atribut panjang rambut tidak akan digunakan karena bukan bagian dari wajah. Model klasifikasi *gender* yang dibuat dengan menggunakan data wajah, diharapkan dapat bermanfaat untuk mengetahui *gender* seseorang tanpa bertanya secara langsung karena *gender* adalah hal yang sensitif bagi sebagian orang. Selain itu, hasil penelitian ini dapat digunakan sebagai salah satu acuan dalam proses pengambilan keputusan pada berbagai situasi seperti rekrutmen pegawai, proses seleksi model kecantikan dan pemeriksaan *gender* untuk atlet olimpiade.

2. METODE PENELITIAN

Metode utama pada penelitian ini mengikuti proses *KDD* (*Knowledge Discovery in Database*). Tahapannya dapat dilihat pada Gambar 2.

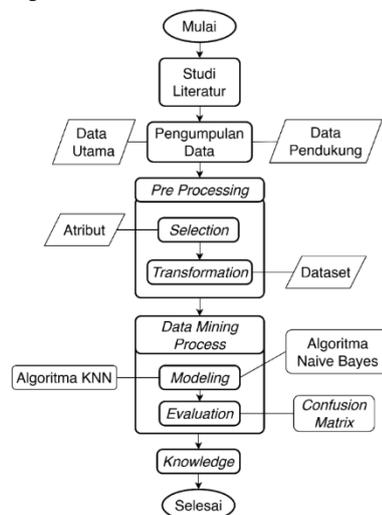
2.1. Studi Literatur

Studi literatur dimulai dengan mengumpulkan sumber literatur yang relevan dengan klasifikasi *gender* dari berbagai sumber yaitu buku, jurnal, artikel dan berita. Informasi yang terkumpul kemudian dianalisis mendalam mulai dari masalah, solusi, metode hingga hasil penelitian. Tujuannya untuk mengidentifikasi celah pengetahuan dari penelitian yang sudah ada.

2.2. Pengumpulan Data

Data yang dikumpulkan terdiri dari data utama dan data pendukung. Data utama merupakan *dataset*

bagian-bagian wajah yang berisi 5001 *record* dengan 7 atribut. Sedangkan untuk data pendukung diperoleh dari situs berita seperti *Sindo News* dan penyedia data statistik seperti *katadata*.



Gambar 2. Metode Penelitian

2.3. Selection

Data yang diperoleh, belum tentu bisa digunakan. Penentuan atribut apa saja yang sesuai dengan tujuan penelitian dilakukan pada tahap ini. Selain itu, data yang terduplikasi akan dibuang, sedangkan untuk data yang salah diperbaiki.

2.4. Transformation

Data yang sudah melalui tahap *selection* diubah formatnya agar dapat digunakan untuk proses *modelling*. Biasanya format *file* dapat dipengaruhi oleh aplikasi yang digunakan sebagai alat untuk mengolah datanya. Penelitian ini akan menggunakan aplikasi *Orange Data Mining*. Salah satu format *file* yang didukung adalah *CSV*.

2.5. Modeling

Tahap pembuatan model dilakukan menggunakan algoritma *Naïve Bayes* dan *K-NN* (*K Nearest Neighbor*). Uji coba dilakukan dengan membagi data menjadi dua yaitu data *training* dan data *testing* dengan perbandingan 80:20.

2.6. Evaluation

Evaluation dilakukan untuk menentukan algoritma mana yang memiliki nilai akurasi yang lebih tinggi. Cara yang akan digunakan pada penelitian ini adalah *Cross Validation* dan *Confusion Matrix*. Kemudian skor nya dilihat dari *AUC* (*Area Under Curve*).

2.7. Knowledge

Knowledge yang ingin diperoleh adalah jawaban dari rumusan masalah. *Knowledge* yang dapat diperoleh adalah kemampuan data ukuran dan bentuk wajah serta algoritma *Naïve Bayes* dan *K-NN* dalam klasifikasi *gender*. Model yang dihasilkan

dapat bermanfaat untuk mengetahui *gender* seseorang tanpa bertanya secara langsung karena *gender* adalah hal yang sensitif bagi sebagian orang. Selain itu, hasil penelitian dapat digunakan sebagai salah satu acuan dalam pengambilan keputusan pada berbagai situasi seperti rekrutmen pegawai, proses seleksi model kecantikan dan pemeriksaan *gender* untuk atlet olimpiade.

3. HASIL DAN PEMBAHASAN

3.1. Studi Literatur

Studi literatur berfokus pada penelitian sejenis yang pernah dilakukan sebelumnya tentang klasifikasi *gender* dengan tipe data teks. Studi literatur dilakukan dengan cara *literature review* untuk mengetahui masalah, metode, hasil, kekurangan serta kelebihan setiap penelitian. *Literature review* dilakukan untuk melihat peluang sehingga dapat membantu menentukan rumusan masalah serta tujuan penelitian. Berdasarkan *literature review* dapat diketahui bahwa data yang digunakan untuk klasifikasi *gender* berbeda-beda. Contohnya seperti *tweet* di *Twitter* (Hussein et al., 2019) (Wani et al., 2019) (Onikoyi et al., 2023) (Roy et al., 2021) komentar di *Facebook* atau *Instagram* (Goenawan et al., 2019), percakapan di *chat room* (Kucukyilmaz et al., 2020) bahkan hingga ukuran badan (Tabassum et al., 2023).

Peluang penelitian yang dikembangkan adalah analisis permasalahan yang menjadi motivasi penelitian. Fenomena yang dilihat tidak hanya bersumber dari penelitian sebelumnya, namun diperkuat dengan kejadian nyata tentang *gender* dari sumber berita. Selain itu, penelitian ini juga memperbaiki kekurangan pada penelitian yang dilakukan oleh (Prayogo, 2022), yaitu pemilihan atribut yang kurang tepat karena memasukan atribut rambut dalam klasifikasi wajah.

3.2. Pengumpulan Data

Data yang dikumpulkan pada penelitian ini dibagi menjadi 2, yaitu data utama dan data pendukung.

3.2.1. Data Utama

Data utama adalah data yang digunakan untuk membuat model klasifikasi. Penelitian ini menggunakan data wajah untuk mengetahui *gender* seseorang. Sumber datanya di diperoleh (Issaden, 2020). Atribut wajah yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Atribut yang Digunakan

No	Atribut	Keterangan
1	<i>Long hair</i> (Panjang Rambut)	Panjang atau tidak
2	<i>Forehead width</i> (Lebar dahi)	Satuan cm
3	<i>Forehead height</i> (Tinggi dahi)	Satuan cm
4	<i>Nose Width</i> (Lebar hidung)	Lebar atau tidak
5	<i>Nose long</i> (Panjang hidung)	Panjang atau tidak
6	<i>Lips thin</i> (Bibir tipis)	Tipis atau tidak

No	Atribut	Keterangan
7	<i>Distance nose to lip</i> (Jarak hidung ke bibir)	Jauh atau tidak
8	<i>Gender</i>	Laki-laki atau perempuan

3.2.2. Data Pendukung

Selain data utama, sumber data lain seperti situs Katadata dan *Sindo News* juga digunakan. Data pendukung yang diperoleh adalah jumlah *gender* di negara Thailand yang mencapai 18 *gender*, Foto yang bisa dipalsukan melalui aplikasi *FaceApp* dan negara dengan jumlah prosedur operasi terbanyak di dunia pada tahun 2020. Jumlah klasifikasi *gender* yang banyak menyebabkan identifikasi *gender* menjadi sulit karena harus mengenal berbagai macam *gender*. Foto yang diedit dengan aplikasi *FaceApp* dapat menyulitkan identifikasi karena kemampuannya mengubah foto laki-laki menjadi perempuan maupun sebaliknya. Jumlah prosedur operasi yang hanya berjumlah 1,5 juta di tahun 2020 memperkuat bahwa data ukuran, fisik dan bentuk wajah dengan tipe data teks masih relevan digunakan untuk klasifikasi *gender*.

3.3. Selection

Selection dilakukan dengan menghapus atribut yang tidak sesuai dengan tujuan penelitian. Atribut yang dihapus pada penelitian ini adalah panjang rambut, karena tujuan pada penelitian ini adalah klasifikasi *gender* dengan menggunakan data wajah, sedangkan rambut adalah bagian dari kepala.

3.4. Transformation

Tahap *transformation* dilakukan pada atribut berikut :

1. *Nose width* : Atribut ini diisi dengan angka 0 untuk menyatakan tidak lebar dan 1 untuk menyatakan lebar.
2. *Nose long* : Atribut ini diisi dengan angka 0 untuk menyatakan tidak panjang dan 1 menyatakan panjang.
3. *Lips thin* : Atribut ini diisi dengan angka 0 untuk menyatakan tebal dan 1 menyatakan tipis
4. *Distance nose to lip* : Atribut ini diisi dengan angka 0 untuk menyatakan dekat dan 1 menyatakan jarak jauh.

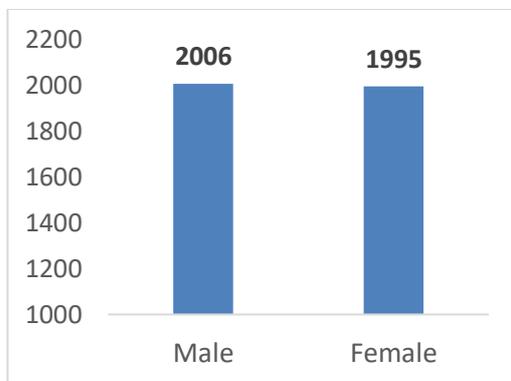
Namun, *transformation* tidak dilakukan pada format *file*, karena aplikasi *Orange Data Mining* sudah mendukung format CSV.

3.5. Modelling

Pemodelan pada penelitian ini menggunakan dua algoritma yaitu *Naïve Bayes* dan K-NN. Pemilihan algoritma *Naïve Bayes* dan *dataset* didasarkan pada penelitian yang dilakukan oleh (Prayogo, 2022). Berdasarkan data yang diperoleh, Algoritma *Naïve Bayes* dipilih karena atribut yang digunakan tidak saling bergantung satu sama lain (*independen*) dan dapat menghasilkan akurasi sebesar 0,85 (Prayogo, 2022), sedangkan algoritma K-NN

dipilih karena nilai dari setiap atribut bisa dikategorikan.

Penentuan jumlah *data training* dan *data testing* didasarkan pada penelitian yang dilakukan oleh (Prayogo, 2022) dengan perbandingan 80:20. *Record* pada *dataset* berjumlah 5001 data. Berdasarkan perbandingan tersebut, 4001 *record* akan menjadi *data training* dan 1000 *record data testing*. Penentuan data mana yang menjadi sampel testing menggunakan sistematis sampling (Sugiyono, 2017). *Data training* dipastikan harus bisa mengenali *gender* laki-laki dan perempuan. Visualisasi sebaran data testing berdasarkan gender dapat dilihat pada Gambar 3. Visualisasi data pada Gambar 3 menunjukkan, *data training* dapat dipakai karena jumlah *record* yang memiliki label *male* (laki-laki) dan *female* (perempuan) selisihnya tidak berbeda jauh.



Gambar 3. Sebaran Gender pada Data Training

1. Naïve Bayes

Pemodelan *Naïve Bayes* dilakukan dengan cara menghitung probabilitas (kemungkinan) dari setiap kriteria. Tahapan dalam pembuatan model dapat dilihat pada Gambar 4.

Berdasarkan hasil klasifikasi algoritma *Naïve Bayes* dapat diketahui ciri-ciri orang yang diklasifikasi sebagai laki-laki yaitu : memiliki bibir yang tipis, jarak antara bibir dan hidung jauh, memiliki lebar dahi lebih dari 14,4 cm memiliki hidung yang panjang dan lebar serta memiliki tinggi dahi lebih dari 6 cm. Sedangkan orang yang di klasifikasikan sebagai perempuan memiliki ciri-ciri : memiliki bibir tebal, jarak antara bibir dan hidung yang dekat, memiliki lebar dahi kurang atau sama dengan 14,4 cm, memiliki hidung yang tidak lebar dan pendek serta tinggi dahinya kurang dari atau sama dengan 6 cm.

2. K-NN

Algoritma K-NN sangat sensitif terhadap penentuan nilai k, karena nilai tersebut akan mempunyai dampak atau pengaruh yang sangat kuat terhadap hasil klasifikasi. Nilai k yang terlalu besar

akan membuat klasifikasi semakin kabur (Fitri Ayuning Tyas et al., 2024) (Putri et al., 2022). Secara keseluruhan, disarankan untuk memilih nilai k berupa angka ganjil untuk menghindari ikatan dalam klasifikasi (ADMINLP2M, 2023). Jika nilai k yang digunakan adalah satu akan mengakibatkan hasil klasifikasi terasa kaku (Windanu et al., 2024).

Berdasarkan penjelasan tersebut, penelitian ini mengambil nilai k dengan nilai ganjil, yaitu 3, 5, 7, 9 dan 11 untuk membantu menemukan nilai k yang optimal untuk klasifikasi. Tahapan dalam pembuatan model dapat dilihat pada Gambar 5.

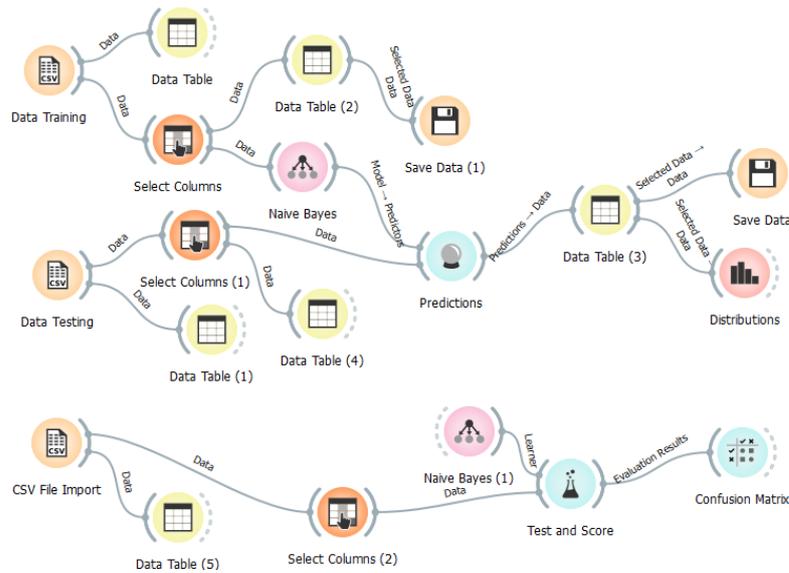
Secara keseluruhan, hasil klasifikasi model K-NN dengan nilai k yang berbeda menghasilkan ciri yang hampir sama. Orang yang diklasifikasi sebagai laki-laki memiliki ciri-ciri : jarak antara bibir dan hidung yang jauh, tinggi dahi lebih dari 6,6 cm, bibir tipis, memiliki hidung panjang dan lebar. Sedangkan orang yang di klasifikasikan sebagai perempuan memiliki ciri-ciri : jarak antara bibir dan hidung dekat, tinggi dahi kurang dari atau sama dengan 6,6 cm, bibir tebal, memiliki hidung yang pendek dan tidak lebar. Perbedaan hasil klasifikasi K-NN berdasarkan nilai k hanya terdapat pada lebar dahi. Ukuran lebar dahi yang dihasilkan masing-masing berdasarkan nilai k dapat dilihat pada Tabel 2.

Tabel 2. Ukuran Lebar Dahi berdasarkan Klasifikasi

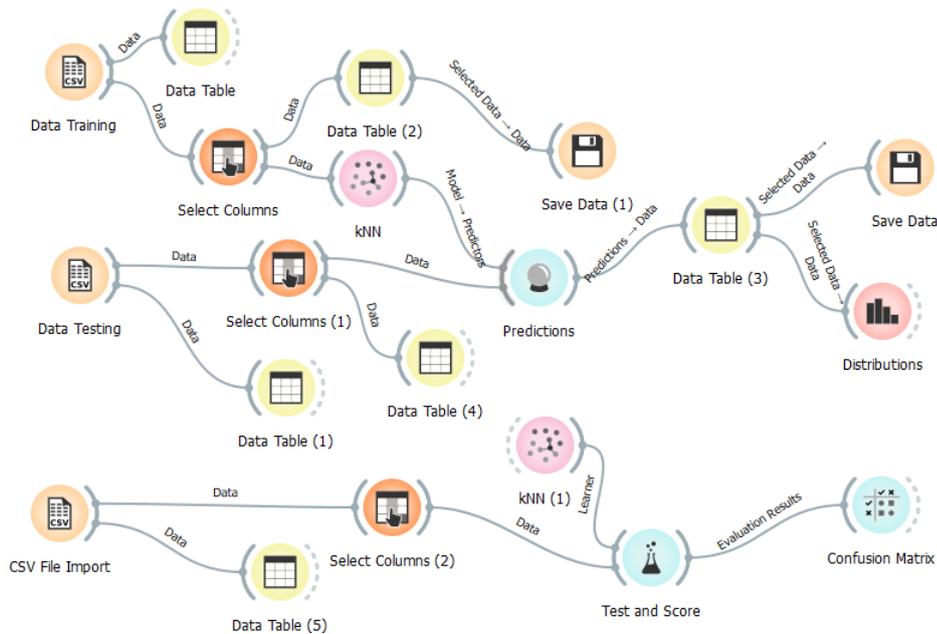
Nilai k	Lebar Dahi	
	Laki-laki	Perempuan
3	Lebih dari 14,4 cm	Kurang dari atau sama dengan 14,4 cm
5	Lebih dari 14,5 cm	Kurang dari atau sama dengan 14,5 cm
7	Lebih dari 14,5 cm	Kurang dari atau sama dengan 14,5 cm
9	Lebih dari 14,4 cm	Kurang dari atau sama dengan 14,4 cm
11	Lebih dari 14,5 cm	Kurang dari atau sama dengan 14,5 cm

3.6. Evaluation

Tahap *Evaluation* dilakukan untuk mengetahui model mana yang paling baik dalam klasifikasi *gender*. Perbandingannya dilihat dengan cara menganalisis nilai AUC yang dihasilkan oleh setiap model. Nilai AUC dipilih karena dapat digunakan untuk klasifikasi biner (dalam penelitian ini laki-laki dan perempuan) dan bisa digunakan untuk membandingkan kinerja model yang berbeda. Tujuannya untuk mengukur secara keseluruhan seberapa baik model dapat membedakan antara laki-laki dan perempuan. Sedangkan *Confusion Matrix* digunakan untuk memberi gambaran yang lebih rinci tentang kuantitas data yang bisa di klasifikasi dengan benar dan salah untuk setiap model.



Gambar 4. Model Naive Bayes pada Orange



Gambar 5. Model K-NN pada Orange (Ganti)

Confusion Matrix adalah metode yang digunakan untuk mengevaluasi performa klasifikasi dalam data mining. Representasi sederhana dari Confusion Matrix dalam hasil klasifikasi diperlihatkan dalam Gambar 6.

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Gambar 6. Confusion Matrix
Sumber : (Kurniawan & Irsyad, 2022)

Confusion Matrix pada Gambar 6 memberikan gambaran tentang akurasi dari prediksi positif yang benar, prediksi positif yang salah, prediksi negatif

yang benar, dan prediksi negatif yang salah. Akurasi dihitung dari semua prediksi yang benar, baik yang positif maupun negatif. Semakin tinggi nilai akurasi, semakin baik kualitas model yang dihasilkan (Widayati et al., 2021).

AUC atau Area Under the Curve adalah sebuah metrik yang digunakan untuk mengevaluasi performa model klasifikasi biner, dengan nilai yang berkisar antara 0 dan 1. AUC menunjukkan kemampuan model dalam membedakan antara dua kelas (misalnya positif dan negatif) di seluruh rentang ambang batas yang mungkin. Semakin tinggi nilai AUC, semakin baik kemampuan model dalam membedakan kelas-kelas tersebut (Carrington et al., 2023).

Nilai AUC memiliki beberapa keunggulan. Pertama, AUC tidak bergantung pada ambang batas tertentu, sehingga memungkinkan model dievaluasi

secara konsisten di berbagai ambang batas yang berbeda. Ini sangat bermanfaat ketika bekerja dengan dataset yang tidak seimbang (misalnya, kasus deteksi penipuan yang jarang terjadi) karena metrik seperti akurasi bisa menyesatkan dengan menunjukkan kinerja yang tinggi hanya karena jumlah kelas mayoritas lebih banyak. AUC memberikan pandangan yang lebih seimbang dengan menilai kinerja model berdasarkan *true positive rate* (sensitivitas) dan *false positive rate* (spesifisitas) pada berbagai ambang batas (Carrington et al., 2023)

1. *Naïve Bayes*

Pengujian model dilakukan dengan metode *Cross Validation*. Hasil pengujian kinerja dengan metode *Cross Validation* pada model *Naïve Bayes* dapat dilihat pada Gambar 7, Gambar 8 dan Tabel 3.

		Predicted		Σ
		Female	Male	
Actual	Female	2447	54	2501
	Male	90	2410	2500
Σ		2537	2464	5001

Gambar 2. *Confusion Matrix Model Naïve Bayes dengan Cross Validation*

		Predicted		Σ
		Female	Male	
Actual	Female	96.5 %	2.2 %	2501
	Male	3.5 %	97.8 %	2500
Σ		2537	2464	5001

Gambar 3. *Confusion Matrix Model Naïve Bayes dengan Cross Validation dalam Persen*

Berdasarkan hasil evaluasi pada *Confusion Matrix* di Gambar 8, kegagalan untuk memprediksi *gender* laki-laki dan perempuan kurang dari 4 persen. Kegagalan tersebut bisa terjadi karena ada orang yang memiliki data yang tidak biasa pada atributnya sehingga model *Naïve Bayes* salah dalam mengklasifikasi *gender*. Contohnya pada atribut *lips thin* (bibir tipis) sebagian besar orang yang memiliki bibir tebal di klasifikasikan sebagai perempuan, padahal kenyataannya, ada juga laki-laki yang memiliki bibir yang tebal. Hasil pengujiannya dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pengujian Algoritma *Naïve Bayes* dengan *Cross Validation*

Model	Number of Folds	AUC
Naïve	2	0,996
Bayes	3	0,996
	5	0,996
	10	0,996
	20	0,996

Tabel 3 menunjukkan model klasifikasi *gender* dengan algoritma *Naïve bayes* memiliki nilai AUC yang stabil di angka 0,996. Hasil tersebut bahkan lebih tinggi dibanding dengan penelitian yang dilakukan oleh (Prayogo, 2022) yang hanya

menghasilkan akurasi sebesar 0,85. Artinya model yang dihasilkan sangat baik. Model dapat membedakan dengan benar 99,6 persen dari sampel dengan *gender* laki-laki dan 99,6 persen dari sampel perempuan. Nilai AUC yang stabil pada algoritma *Naïve Bayes* terjadi karena atribut yang digunakan tidak memiliki keterkaitan satu sama lain. Atribut yang tidak terkait dalam penelitian ini misalnya antara panjang hidung dan lebar hidung, bibir tipis dan jarak bibir ke hidung. Orang yang memiliki hidung yang panjang, belum tentu memiliki hidung yang lebar, begitu juga sebaliknya. Orang yang memiliki bibir tipis, belum tentu memiliki jarak bibir ke hidung yang jauh. Atribut yang tidak saling terkait juga menyebabkan model *Naïve Bayes* memiliki nilai AUC yang tinggi.

2. K-NN

Seperti pada model *Naïve Bayes*, pengujian model K-NN juga dilakukan dengan metode *Cross Validation*. Hasil pengujian kinerja dengan metode *Cross Validation* pada model K-NN dapat dilihat pada Gambar 9, Gambar 10 dan Tabel 4.

		Predicted		Σ
		Female	Male	
Actual	Female	2459	54	2513
	Male	87	2401	2488
Σ		2546	2455	5001

Gambar 4. *Confusion Matrix Model K-NN dengan Cross Validation*

		Predicted		Σ
		Female	Male	
Actual	Female	96.6 %	2.2 %	2513
	Male	3.4 %	97.8 %	2488
Σ		2546	2455	5001

Gambar 5. *Confusion Matrix Model K-NN dengan Cross Validation dalam Persen*

Berdasarkan hasil evaluasi pada *Confusion Matrix* pada Gambar 10, kegagalan untuk memprediksi laki-laki dan perempuan kurang dari 4 persen. Sama seperti model *Naïve Bayes*, kegagalan tersebut bisa terjadi karena ada *gender* yang memiliki data yang tidak biasa pada atributnya sehingga model K-NN salah dalam mengklasifikasi *gender*. Contohnya pada atribut *lips thin* (bibir tipis), sebagian besar orang yang memiliki bibir tebal di klasifikasikan sebagai perempuan, padahal kenyataannya, ada juga laki-laki yang memiliki bibir yang tebal. Hasil pengujiannya dapat dilihat pada Tabel 4.

Tabel 4 menunjukkan, nilai k yang menghasilkan akurasi tinggi adalah 11. Hal tersebut dapat dilihat dari hasil uji yang memperlihatkan dengan nilai k=11 dan nilai *fold* yang berbeda-beda,

mampu menghasilkan nilai AUC yang paling tinggi yaitu 0,992. Artinya model yang dihasilkan juga sangat baik. Model dapat membedakan dengan benar 99,2 persen dari sampel dengan *gender* laki-laki dan 99,2 persen dari sampel perempuan.

Tabel 4. Hasil Pengujian Algoritma K-NN dengan *Cross Validation*

Model	k	Number of Folds	AUC
K-NN	3	2	0,968
		3	0,971
		5	0,970
		10	0,971
		20	0,971
	5	2	0,968
		3	0,969
		5	0,970
		10	0,971
		20	0,971
	7	2	0,968
		3	0,969
		5	0,970
		10	0,971
		20	0,971
	9	2	0,969
		3	0,971
		5	0,970
		10	0,971
		20	0,972
11	2	0,970	
	3	0,972	
	5	0,971	
	10	0,971	
	20	0,972	

Nilai akurasi tinggi pada model K-NN dengan $k=11$ terjadi karena algoritma tersebut menggunakan paling banyak tetangga untuk dipelajari kemiripannya, sehingga lebih akurat dalam mengklasifikasikan data. Penyebab lain tingginya nilai AUC pada model K-NN di penelitian ini adalah atribut yang relevan dan jumlah data yang cukup. Klasifikasi *gender* pada penelitian ini didasarkan pada wajah, sehingga atribut yang digunakan hanya yang berada di bagian wajah. Data yang cukup dapat dilihat dari jumlah data *training* yang dimiliki sudah mewakili masing-masing *gender*. Data *training* yang digunakan berisi 2006 *gender* laki-laki dan 1995 data *gender* perempuan. Jumlah ini cukup mewakili pola data *gender* laki-laki dan *gender* perempuan karena perbedaannya tidak terlalu jauh.

3.7. Knowledge

Knowledge (pengetahuan) yang dihasilkan dari penelitian ini adalah kondisi fisik wajah dapat digunakan untuk mengidentifikasi perbedaan *gender* antara laki-laki dan perempuan. Keunggulan model klasifikasi wajah dengan hasil pengukuran secara fisik adalah akurasi yang lebih tinggi dari klasifikasi yang hanya mengandalkan foto karena foto sangat rentan untuk dimodifikasi. Model pada penelitian ini bisa digunakan pada berbagai profesi seperti mahasiswa, pengusaha, HRD, panitia model kecantikan, hingga panitia olimpiade. Bagi mahasiswa, penelitian ini dapat digunakan sebagai

salah satu acuan untuk penelitian selanjutnya tentang bagaimana data wajah bisa digunakan untuk klasifikasi *gender*, kinerja algoritma *Naïve bayes* dan K-NN dalam melakukan klasifikasi *gender* serta perbandingan akurasi dari kedua algoritma tersebut.

Bagi pengusaha, model pada penelitian ini dapat digunakan dalam proses seleksi pada saat rekrutmen pegawai untuk pekerjaan yang memiliki syarat *gender* tertentu, melakukan demografi konsumen berdasarkan *gender* untuk mengetahui jenis kelamin sebenarnya tanpa harus bertanya tentang *gendernya* secara langsung kepada pelamar kerja.

Pada bidang olahraga dan hiburan, model klasifikasi *gender* juga bisa digunakan pada proses seleksi peserta lomba ajang kecantikan untuk memastikan apakah peserta benar-benar perempuan atau tidak. Selain itu, model ini juga bisa digunakan dalam pemeriksaan *gender* pada atlet olimpiade. Tujuannya untuk mencegah kecurangan, karena ada olahraga yang hanya bisa diikuti oleh *gender* tertentu saja seperti voli putri yang hanya bisa diikuti pemain perempuan dan voli putra yang hanya bisa diikuti oleh laki-laki.

Terlepas dari manfaat yang diperoleh dari adanya model klasifikasi *gender*, perlu diperhatikan bahwa model yang dihasilkan tidak sepenuhnya akurat. Validasi dari orang yang ahli pada bidangnya yaitu dokter kecantikan dan dokter bedah tetap diperlukan. Tujuannya untuk meminimalisir kekeliruan yang disebabkan oleh model klasifikasi *gender* yang digunakan.

4. KESIMPULAN

Berdasarkan pengujian akurasi model yang dilakukan, atribut *forehead width* (lebar dahi), *forehead height* (tinggi dahi), *nose width* (lebar hidung), *nose long* (panjang hidung), *lips thin* (bibir tipis) dan *distance nos to lip* (jarak hidung ke bibir) dapat digunakan untuk klasifikasi *gender* karena menghasilkan akurasi yang sangat baik. Jika melihat dari AUC, kedua algoritma memiliki akurasi yang baik. Namun, algoritma *Naïve Bayes* lebih direkomendasikan karena memiliki nilai AUC yang lebih tinggi yaitu 0,996, sedangkan pada algoritma K-NN nilai AUC tertingginya hanya 0,992 dengan nilai $k=11$. Selain itu, model yang dihasilkan juga lebih akurat, karena data yang digunakan adalah hasil pengukuran dan bentuk wajah secara fisik, bukan berupa foto wajah yang dapat dimodifikasi.

DAFTAR PUSTAKA

- ADMINLP2M. 2023, February 16. *Algoritma K-Nearest Neighbors (KNN) – Pengertian dan Penerapan*. <https://lp2m.uma.ac.id/2023/02/16/algoritma-k-nearest-neighbors-knn-pengertian-dan-penerapan/> (Diakses tanggal 23 November 2023)
- AJIJAH, N., & KURNIAWAN, A. 2023. Klasifikasi

- Teks Mining Terhadap Analisa Isu Kegiatan Tenaga Lapangan Menggunakan Algoritma K-Nearest Neighbor (KNN). In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 7, Issue 1). <http://dx.doi.org/10.30645/j-sakti.v7i1.589>
- APRILIA, T. 2024. Klasifikasi Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor dan Metode Naive Bayes. *SATESI (Jurnal Sains Teknologi Dan Sistem Informasi)*, 4(2), 156–163. <https://doi.org/10.54259/satesi.v4i2.3167>
- ARMANDHANI, R., CAHYA WIHANDIKA, R., & RAHMAN, M. A. 2019. *Klasifikasi Gender berbasis Wajah menggunakan Metode Local Binary Pattern dan Random KNN* (Vol. 3, Issue 8). <http://j-ptiik.ub.ac.id>
- AZIZAH, N. 2021. Aliran Feminis dan Teori Kesetaraan Gender dalam Hukum. In *SPECTRUM: Journal of Gender and Children Studies* (Vol. 1, Issue 1). <http://journal.iain-manado.ac.id/index.php/SPECTRUM> <https://doi.org/10.30984/spectrum.v1i1.163>
- BAHTIAR, R. 2023. Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor. In *Jurnal Informatika MULTI* (Vol. 1, Issue 3). <https://jurnal.publikasitecno.id/index.php/jim203>
- BUULOLO, E. 2020. *Data Mining untuk Perguruan Tinggi*. Deepublish .
- CARRINGTON, A. M., MANUEL, D. G., FIEGUTH, P. W., RAMSAY, T., OSMANI, V., WERNLY, B., BENNETT, C., HAWKEN, S., MAGWOOD, O., SHEIKH, Y., MCINNES, M., & HOLZINGER, A. 2023. Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 329–341. <https://doi.org/10.1109/TPAMI.2022.3145392>
- DANAR. 2019. *Tragis! Aplikasi FaceApp Abal-Abal Tipu Ribuan Pengguna di Indonesia - Krjogja*. <https://www.krjogja.com/teknologi/1242545082/tragis-aplikasi-faceapp-abalabal-tipu-ribuan-pengguna-di-indonesia> (Diakses tanggal 21 September 2023)
- DEVIAN, SABRINA, P. N., & KOMARUDIN, A. 2024. Prediksi Penyakit Diabetes Dengan Metode K-Nearest Neighbor (KNN) Dan Seleksi Fitur Information Gain. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(6), 11320–11326. <https://doi.org/10.36040/jati.v8i6.11364>
- DEVITO, D., CAHYA WIHANDIKA, R., & WIDODO, A. W. 2019. *Ekstraksi Ciri Untuk Klasifikasi Gender Berbasis Citra Wajah Menggunakan Metode Histogram of Oriented Gradients* (Vol. 3, Issue 8). <http://j-ptiik.ub.ac.id>
- DW. 2022. *AS Perkenalkan Paspor “X” yang Netral Gender*. <https://www.dw.com/id/as-perkenalkan-paspor-x-yang-netral-gender/a-61325255> (Diakses tanggal 11 Oktober 2023)
- FANSYURI, M., & YUNITA, D. 2024. Analisa Citra Wajah Untuk Identifikasi Klasifikasi Jenis Kelamin Menggunakan Algoritma Naive Bayes. *LOGIC: Jurnal Ilmu Komputer Dan Pendidikan*, 2(3), 594–606. <https://journal.mediapublikasi.id/index.php/logic>
- FAUZIAH, S. 2023. *18 Gender di Thailand yang Belum Banyak Diketahui, Apa Saja?* <https://lifestyle.sindonews.com/read/1024453/166/18-gender-di-thailand-yang-belum-banyak-diketahui-apa-saja-1676534541> (Diakses tanggal 21 September 2023)
- FEBRINAMAS, D. R., HIDAYATI, R., & NIRMALA, I. 2023. Klasifikasi Buah Pinang Berdasarkan Data Sensor Menggunakan Metode K-Nearest Neighbor Berbasis Web. *Journal of Computer System and Informatics (JoSYC)*, 4(4). <https://doi.org/10.47065/josyc.v4i4.3805>
- FITRI AYUNING TYAS, MAHDA NURAYUNI, & HIDAYATUR RAKHMAWATI. 2024. Optimasi Algoritma K-Nearest Neighbors Berdasarkan Perbandingan Analisis Outlier (Berbasis Jarak, Kepadatan, LOF). *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 13(2), 108–115. <https://doi.org/10.22146/jnteti.v13i2.9579>
- GOENAWAN, N. R., CHANRICO, W., SUHARTONO, D., & PURNOMO, F. 2019. Gender demography classification on instagram based on user’s comments section. *Procedia Computer Science*, 157, 64–71. <https://doi.org/10.1016/j.procs.2019.08.142>
- HUSSEIN, S., FAROUK, M., & HEMAYED, E. S. 2019. Gender identification of egyptian dialect in twitter. *Egyptian Informatics Journal*, 20(2), 109–116. <https://doi.org/10.1016/j.eij.2018.12.002>
- ISSADEN, J. 2020. *Gender Classification Dataset /Kaggle*. <https://www.kaggle.com/datasets/elakiricoder/gender-classification-dataset> (Diakses tanggal 30 Juni 2023)
- KUCUKYILMAZ, T., DENIZ, A., & KIZILOZ, H. E. 2020. Boosting gender identification using author preference. *Pattern Recognition Letters*, 140, 245–251. <https://doi.org/10.1016/j.patrec.2020.10.002>
- KURNIAWAN, C., & IRSYAD, H. 2022. Perbandingan Metode K-Nearest Neighbor Dan Naive Bayes Untuk Klasifikasi Gender Berdasarkan Mata. *Jurnal Algoritme*, 2(2). <https://doi.org/10.35957/algoritme.v2i2.2358>

- MUHADI, A., & OCTAVIANO, A. 2023. Penerapan Data Mining Untuk Prediksi Hasil Keuntungan Lelang Mesin X-Ray Tahun 2020 Dengan Metode K-Nearest Neighbor (Studi Kasus : PT.Ramadika Mandiri). In *Jurnal Informatika MULTI* (Vol. 1, Issue 2). <https://jurnal.publikasitecno.id/index.php/jim126>
- ONIKOYI, B., NNAMOKO, N., & KORKONTZELOS, I. 2023. Gender prediction with descriptive textual data using a Machine Learning approach. *Natural Language Processing Journal*, 4, 100018. <https://doi.org/10.1016/j.nlp.2023.100018>
- PRAYOGO, A. A. T. 2022. Penerapan Model Klasifikasi Untuk Prediksi Gender Berdasrakan Wajah Dengan Menggunakan Metode Naïve Bayes. *JURNAL INDUSTRI KREATIF DAN INFORMATIKA SERIES (JIKIS)*, 02(02).
- PUTRI, V. A. P., PRASETIJO, A. B., & ERIDANI, D. 2022. Perbandingan Kinerja Algoritma Naïve Bayes Dan K-Nearest Neighbor (Knn) Untuk Prediksi Harga Rumah. *Jurnal Ilmiah Teknik Elektro*, 24(2), 162–171. <https://ejournal.undip.ac.id/index.php/transmisi>
<https://doi.org/10.14710/transmisi.24.4.162-171>
- ROY, S., MITTAL, M., KIM, T. H., DHAYAL, K., BHATTACHARYYA, D., & PATEL, B. 2021. Demographical gender prediction of Twitter users using big data analytics: an application of decision marketing. *International Journal of Reasoning-Based Intelligent Systems*, 13(2), 41. <https://doi.org/10.1504/ijris.2021.10036801>
- SUGIYONO. 2017. *Statistika untuk Penelitian*. Alfabeta.
- TABASSUM, H., IQBAL, M. M., MAHMOOD, Z., PARVEEN, M., & ULLAH, I. 2023. Gender classification from anthropometric measurement by boosting decision tree: A novel machine learning approach. *Journal of the National Medical Association*, 115(3), 273–282. <https://doi.org/10.1016/j.jnma.2022.12.005>
- UTAMI, M., & AYUMI, V. 2024. Prediksi Penjualan Produk Terlaris Menggunakan Algoritma K-Nearest Neighbour (KNN). *JCOSIS (Journal Computer Science and Information Syetem)*, 43–47. <https://doi.org/10.61567/jcosis.v1i2.214>
- WAHYU, M., SANTOSO, B., CAHYA WIHANDIKA, R., & RAHMAN, M. A. 2019. *Ekstraksi Ciri untuk Klasifikasi Jenis Kelamin berbasis Citra Wajah menggunakan Metode Compass Local Binary Patterns* (Vol. 3, Issue 11). <http://j-ptiik.ub.ac.id>
- WANI, M. A., BOURS, P., AGRAWAL, N., & JABIN, S. 2019. Emotion-based Mining for Gender Prediction in Online Social Networks. *ACM International Conference on Machine Learning and Data Science (ICMLDS)-2019*, 56–60. <https://doi.org/10.1145/1234567890>
- WIDAYATI, Y. T., PRIHATI, Y., & WIDJAJA, S. 2021. Pelanggan Mnc Play Kota Semarang. *TRANSFORMTIKA*, 18(2), 161–172. <http://dx.doi.org/10.26623/transformatika.v18i2.2541>
- WINDANU, L. S., WIGUNA, A. S., & BUDIANTO, A. E. 2024. Optimasi Metode K-Nearest Neighbor Berbasis Particle Swarm Optimization Untuk Analisis Sentimen Pemilihan Presiden Indonesia Tahun 2024-2029. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 35–42. <https://doi.org/10.36040/jati.v8i1.6874>

Halaman ini sengaja dikosongkan