

## KLASIFIKASI JENIS KELAMIN BERDASARKAN CIRI FISIK MENGUNAKAN ALGORITMA NEURAL NETWORK

<sup>1</sup>Agung M. Rizki, <sup>2</sup>Gusti E. Yuliasuti, <sup>3</sup>Eva Y. Puspaningrum <sup>4</sup>Afina Lina Nurlaili  
<sup>1,3,4</sup>UPN "Veteran" Jawa Timur, <sup>2</sup>Institut Teknologi Adhi Tama Surabaya  
Email: [agung.mustika.if@upnjatim.ac.id](mailto:agung.mustika.if@upnjatim.ac.id)

**Abstrak.** Pengenalan merupakan sebuah kegiatan yang sulit dilakukan karena setiap manusia memiliki ciri fisik yang tidak sama kecuali pada kasus khusus seperti kembar identik. Untuk mengatasi hal tersebut diperlukan penentuan ciri fisik yang akan digunakan sebagai kriteria pengenalan. Dataset yang digunakan pada penelitian ini adalah data ciri-ciri fisik pada bagian kepala manusia yang berupa panjang rambut, lebar dahi, tinggi dahi, lebar hidung, tinggi hidung, ketebalan bibir, serta jarak hidung dengan bibir. Dengan ciri fisik tersebut dapat diterapkan algoritma Artificial Neural Network (ANN) untuk menentukan jenis kelamin. Algoritma ANN dipilih karena berdasarkan beberapa penelitian sebelumnya terbukti efektif dengan nilai akurasi yang cukup tinggi. Lebih khusus metode multi layer perceptron dipilih untuk mendukung kinerja algoritme ANN. Hasilnya didapatkan nilai precision perempuan sebesar 97% dan laki-laki 96%. Sementara untuk nilai recall dan f1-score masing masing memiliki nilai yang sama yakni perempuan sebesar 96% dan laki-laki 97%. Secara lebih lanjut penelitian ini menghasilkan nilai akurasi rata-rata akurasi di angka 97%.

**Kata Kunci:** Klasifikasi, Jenis Kelamin, Ciri Fisik, Multi Layer, Neural Network

Dengan perkembangan teknologi yang semakin cepat, data biologis manusia sudah banyak digunakan pada perangkat elektronik untuk mendukung kinerja sebuah perangkat. Paling sering dijumpai adalah pengenalan ciri fisik manusia khususnya bagian wajah yang merupakan bagian tubuh yang paling mudah untuk diidentifikasi. Contoh pemanfaatannya adalah fitur pengenalan wajah untuk kunci keamanan perangkat.

Pengenalan merupakan sebuah kegiatan yang sulit dilakukan karena setiap manusia memiliki ciri fisik yang tidak sama kecuali pada kasus khusus seperti kembar identic [7]. Untuk mengatasi hal tersebut diperlukan penentuan ciri fisik yang akan digunakan sebagai kriteria pengenalan.

Dataset yang digunakan pada penelitian ini adalah data ciri-ciri fisik pada bagian kepala manusia yang berupa Panjang rambut, lebar dahi, tinggi dahi, lebar hidung, tinggi hidung, ketebalan bibir, serta jarak hidung dengan bibir.[5]

Klasifikasi adalah sebuah proses menentukan arti kesamaan karakteristik pada suatu kelompok atau kelas (class) berdasarkan fitur/kriteria yang ada(Asmara et al., 2018). Klasifikasi data mining dilakukan untuk memperkirakan kelas dari suatu objek yang labelnya belum diketahui. Dalam penelitian ini akan ada 2 kelas yakni laki-laki dan perempuan yang diperoleh dari kriteria yang sudah disebutkan berdasarkan algoritma *Artificial Neural Network* (ANN).

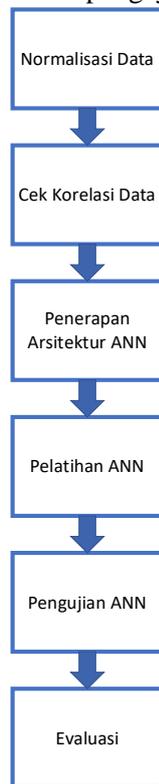
*Artificial Neural Network* (ANN) bisa juga disebut Jaringan saraf tiruan (JST) adalah jaringan dari sekelompok unit pemrosesan kecil yang terinspirasi berdasarkan sistem saraf manusia. ANN merupakan sistem yang bisa mengubah strukturnya secara adaptif untuk memecahkan masalah mengacu pada informasi eksternal maupun internal yang mengalir melalui jaringan tersebut[2].

Secara ringkas, ANN adalah sebuah media pemodelan data statistik yang bersifat non-linier. ANN juga digunakan untuk memodelkan keterkaitan yang kompleks antara *input* dan *output* agar dapat menemukan pola-pola pada data.

Pemilihan algoritma ANN pada penelitian ini mengacu pada performa algoritma ANN dalam beberapa penelitian yang sudah dilakukan untuk melakukan klasifikasi. Pada penelitian komparasi algoritma neural network dan naïve bayes untuk memprediksi penyakit jantung menghasilkan akurasi sebesar 75% [6]. Selanjutnya pada penelitian klasifikasi jenis kelamin wajah bermasker menggunakan algoritma supervised learning menghasilkan rata-rata akurasi sebesar 98% [1]. Dibandingkan dengan penelitian klasifikasi jenis kelamin pada citra wajah menggunakan metode naïve bayes menghasilkan akurasi sebesar 80% [4].

### I. Metodologi

Alur metodologi yang digunakan pada penelitian ini seperti digambarkan pada Gambar 1 yaitu normalisasi data, cek korelasi data, penerapan arsitektur ANN, pelatihan ANN, pengujian ANN dan evaluasi ANN berdasarkan hasil dari pelatihan dan pengujian yang didapat.



Gambar 1. Alur Metodologi

#### Normalisasi Data

Dataset yang digunakan pada penelitian 5001 data yang terdiri dari data ciri fisik pada bagian wajah manusia. orang yang berbeda. Ada 7 ciri fisik yang digunakan yakni Panjang rambut, lebar dahi, tinggi dahi, lebar hidung, tinggi hidung, ketebalan bibir, serta jarak hidung dengan bibir. Dari ciri fisik tersebut selanjutnya difungsikan sebagai kriteria/fitur pada algoritma ANN.

Dataset dibagi menjadi 2 bagian, yaitu 70% dataset pelatihan (3501 data) dan 30% dataset pengujian (1500 data). Dataset yang ada memiliki range nilai yang berbeda. Selanjutnya dilakukan normalisasi untuk mendapatkan range data bernilai 0.1 sampai dengan 0.9 menggunakan rumus:

$$x' = \frac{x}{max} \tag{1}$$

Ket:

- $x'$  = hasil normalisasi
- $x$  = nilai
- $max$  = nilai tertinggi dari data fitur

#### Cek Korelasi Data

Analisis korelasi merupakan teknik analisis yang dipakai untuk mengetahui kedekatan hubungan antara beberapa variabel. Uji korelasi ini akan menunjukkan apakah masing-masing variabel saling mempengaruhi. Walaupun variabel tersebut memiliki keterikatan erat atau berkorelasi, tidak menjamin variabel tersebut saling mempengaruhi. Pada analisis korelasi, output yang dihasilkan hanya dalam rentang -1 sampai 1 dan terbagi menjadi korelasi positif (hasil positif), korelasi negatif (hasil negatif), dan tidak berkorelasi sama sekali (0). Rumus korelasi:

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2 y^2)}}$$

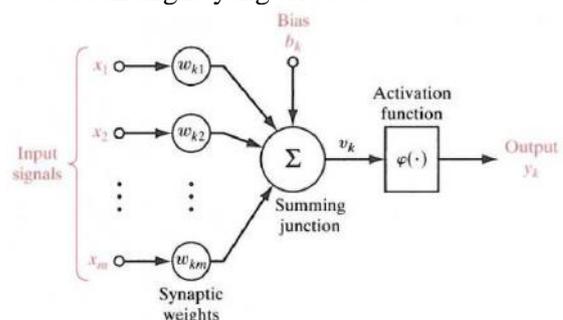
Dimana :

- $r_{xy}$  = korelasi antara variabel x dan y
- $x$  =  $(X_i - \bar{X})$
- $y$  =  $(Y_i - \bar{Y})$

#### Penerapan Arsitektur ANN

Prinsip dari *Artificial Neural Network* (ANN) ditentukan oleh tiga elemen dasar model, yaitu :

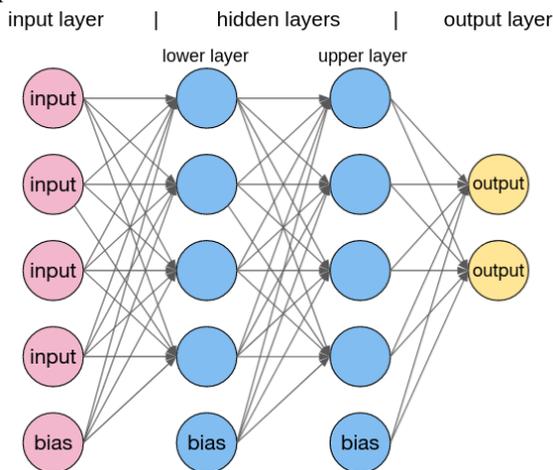
1. Satu set dari sinapsis, atau penghubung yang masing-masing digolongkan oleh bobot atau kekuatannya.
2. Sebuah penambah untuk menjumlahkan sinyal-sinyal input. Ditimbang dari kekuatan sinaptik masing-masing neuron.
3. Sebuah fungsi aktivasi untuk membatasi amplitudo *output* dari *neuron*. Fungsi ini bertujuan membatasi jarak amplitude yang diperbolehkan oleh sinyal *output* menjadi sebuah angka yang terbatas.



Gambar 2. Prinsip dasar JST

Pada penelitian ini terdapat 4 fitur sebagai inputan, 2 hidden layer dengan masing-masing memiliki 12 dan 8 node serta 2 kelas

sebagai output. Sebagai gambaran dapat dilihat pada Gambar 3 berikut:



Gambar 3. Arsitektur ANN  
(An, n.d.)

### Pelatihan ANN

Pada penelitian ini ANN membutuhkan proses pelatihan agar dapat melakukan prediksi kelas suatu data uji baru yang ditemukan. Proses pelatihan ANN menggunakan metode: *Perceptron*, *Backpropagation*, *Self Organizing Mapping*, *Delta*, *Associative Memory*, *Learning Vector Quantization*.

Dari beberapa metode yang telah disebutkan, multi layer perceptron dipilih untuk menyelesaikan permasalahan klasifikasi jenis kelamin ini. *Multi layer perceptron* merupakan ANN turunan dari *perceptron* yang berupa ANN *feedforward* dengan satu atau lebih *hidden layer*.

### Pengujian ANN

Pengujian dilakukan setelah proses pelatihan dataset selesai dilakukan. Hal ini dimaksudkan agar layer perceptron mendapatkan komposisi terbaiknya.

### Evaluasi

Proses evaluasi dilakukan dengan perhitungan akurasi. Pada penelitian ini digunakan proses evaluasi menampilkan sebaran hasil perhitungan melalui matriks. Penampilan matriks menggunakan 2 grafik yang dimana dari masing-masing grafik menampilkan perbandingan hasil perhitungan antara lain *loss* pada pelatihan dan pengujian data dengan *epoch*, akurasi pada pelatihan dengan *epoch*.

Selanjutnya untuk menguji performa digunakan *confusion matrix* yang sering dipakai untuk masalah klasifikasi *machine learning*.

Keluaran *confusion matrix* dapat berupa dua kelas atau lebih. *confusion matrix* adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual.

## II. Hasil dan Pembahasan

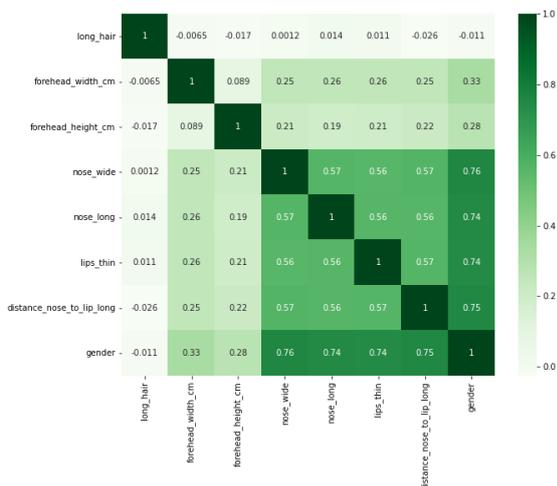
Pada penelitian ini kami menggunakan *google collab* sebagai perangkat untuk menuliskan kode program dan Bahasa pemrograman yang digunakan adalah bahasa pemrograman *python*. Hasil dari penelitian ini ditunjukkan mulai dari proses normalisasi data sampai dengan evaluasi. Berikut data hasil normalisasi:

Tabel 1. Data Hasil Normalisasi

| No   | Panjang rambut | lebar dahi | tinggi dahi | .... | jenis kelamin |
|------|----------------|------------|-------------|------|---------------|
| 0    | 1              | 0.7613     | 0.8592      | ...  | 1             |
| 1    | 0              | 0.9032     | 0.7606      | ...  | 0             |
| 2    | 0              | 0.7613     | 0.8873      | ...  | 1             |
| 3    | 0              | 0.929      | 0.8592      | ...  | 1             |
| ...  | ...            | ...        | ...         | ...  | ...           |
| 4997 | 1              | 0.7677     | 0.7606      | ...  | 0             |
| 4998 | 1              | 0.8323     | 0.8028      | ...  | 0             |
| 4999 | 1              | 0.8516     | 0.8732      | ...  | 0             |
| 5000 | 1              | 0.9935     | 0.7606      | ...  | 1             |

Dari proses normalisasi didapatkan data dengan range nilai 0 sampai 1 pada setiap fitur. Hal ini perlu dilakukan agar setiap fitur memiliki range nilai yang sama sehingga tidak ada salah satu fitur yang dominan karena perbedaan *range* nilai.

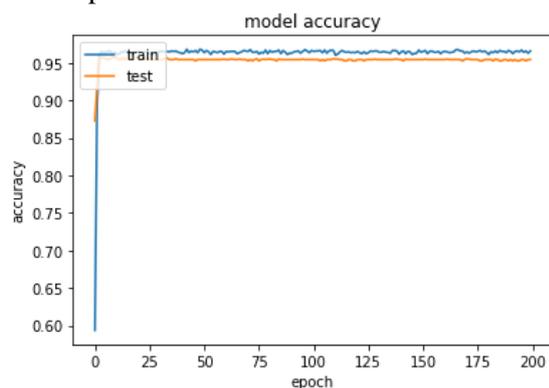
Selanjutnya proses mengecek korelasi data dari setiap fitur. Proses ini akan mengeliminasi fitur yang tidak memiliki kontribusi yang diinginkan dalam menentukan kelas. Berikut adalah hasil dari pengecekan korelasi data dari setiap fitur:



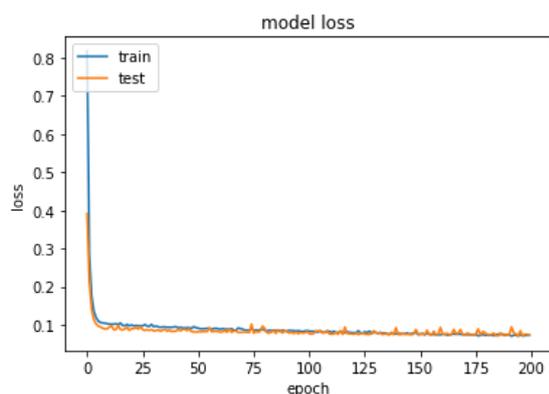
Gambar 4 Hasil Korelasi Data

Dari proses pengecekan korelasi didapatkan 4 fitur yang berpengaruh terhadap kelas jenis kelamin yaitu lebar hidung, tinggi hidung, ketebalan bibir, dan jarak hidung dengan bibir. Sementara 3 fitur lain yakni panjang rambut, lebar dahi dan tinggi dahi tidak digunakan karena memiliki nilai korelasi terhadap kelas jenis kelamin di bawah rata-rata.

Pada proses pelatihan dan pengujian digunakan `variable validation_split=0.5`, `epochs = 200`, `batch_size = 10` dan didapatkan hasil seperti berikut:



Gambar 5 Model Akurasi



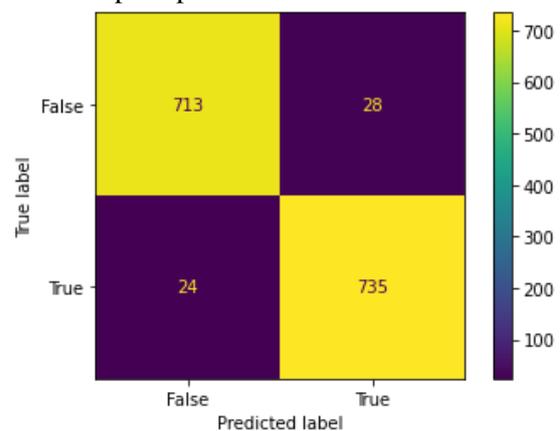
Gambar 6 Model Loss

Pada Gambar 4 menunjukkan grafik model akurasi yang dapat dilihat bahwa nilai akurasi cukup tinggi dari *epoch* awal di angka 90% dan seiring bertambahnya nilai *epoch*, nilai akurasi juga mengalami kenaikan, baik untuk training dataset maupun testing dataset. Meskipun dapat dilihat juga bahwa nilai akurasi dari pengujian data tidak lebih besar daripada nilai akurasi dari pelatihan data.

Sementara pada Gambar 5 menunjukkan nilai loss yang cenderung menurun seiring bertambahnya nilai *epoch*. Nilai *loss* pada pengujian dataset secara umum memiliki kemiripan atau bisa dibilang bersinggungan dengan nilai *loss* dari pelatihan dataset.

Pergerakan nilai akurasi yang semakin meningkat dan nilai *loss* yang cenderung menurun ini menunjukkan bahwa arsitektur ANN berjalan dengan baik dan menghasilkan hasil yang cukup efektif.

Pada tahap terakhir pengujian performa dilakukan dengan menggunakan *confusion matrix* seperti pada Gambar 7 berikut:



Gambar 7 Confusion Matrix

Pada Gambar 7 dapat ditunjukkan hasil confusion matrix sebagai berikut:

- Nilai True Positive (TP) yaitu model memprediksi dengan benar kelas positif (prediksi dan aktual keduanya positif) sejumlah 735 data yang memiliki jenis kelamin laki-laki diprediksi positif oleh model.
- Nilai True Negative (TN) — model memprediksi kelas negatif dengan benar (prediksi dan aktual keduanya negatif) sejumlah 713 orang yang memiliki jenis kelamin perempuan diprediksi secara negatif oleh model.
- Nilai False Positive (FP) model memberikan prediksi yang salah dari kelas negatif (predicted-positive, actual-

negative) sejumlah 28 orang diprediksi berjenis kelamin laki-laki, padahal jenis kelaminnya adalah perempuan.

- Niali False Negative (FN) — model salah memprediksi kelas positif (predicted-negative, actual-positive) sejumlah 24 orang yang berjenis kelamin laki-laki diprediksi perempuan.

Dari penelitian ini juga didapatkan hasil nilai *precision* perempuan sebesar 97% dan laki-laki 96%. Sementara untuk nilai *recall* dan *f1-score* masing masing memiliki nilai yang sama yakni perempuan sebesar 96% dan laki-laki 97%. Secara lebih lanjut penelitian ini menghasilkan nilai akurasi rata-rata akurasi di angka 97%.

### III. Kesimpulan

Penggunaan arsitektur ANN dalam permasalahan klasifikasi jenis kelamin berdasarkan ciri fisik menghasilkan akurasi terbesar sebesar 97% dan juga loss test terkecil dibawah 1,%. Hasil ini tentu saja membuktikan bahwa arsitektur ANN bekerja secara efektif.

Namun tidak berhenti di sini, arsitektur ANN masih dapat dikembangkan menjadi banyak variasi. Sehingga nilai akurasi dan loss yang didapatkan pada penelitian ini masih dapat ditingkatkan lagi.

### IV. Daftar Pustaka

- [1] Adinata, F. D., & Arifin, J. (2022). Klasifikasi Jenis Kelamin Wajah Bermasker Menggunakan Algoritma Supervised Learning. *Jurnal Media Informatika Budidarma*, 6(1), 229. <https://doi.org/10.30865/mib.v6i1.3377>
- [2] Alwi, A. A., Adikara, P. P., & Indriati. (2020). Pengenalan Jenis Kelamin dan Rentang Umur berdasarkan Suara menggunakan Metode Backpropagation Neural Network. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(7), 2083–2093.
- [3] An, S. (n.d.). *Introduction to how an Multilayer Perceptron works but without complicated math*. Retrieved September 12, 2022, from <https://medium.com/codex/introduction-to-how-an-multilayer-perceptron-works-but-without-complicated-math-a423979897ac>
- [4] Asmara, R. A., Andjani, B. S., Rosiani, U. D., & Choirina, P. (2018). Klasifikasi Jenis Kelamin Pada Citra Wajah

Menggunakan Metode Naive Bayes. *Jurnal Informatika Polinema*, 4(3), 212.

<https://doi.org/10.33795/jip.v4i3.209>

- [5] Kaggle. (n.d.). *Gender Classification Dataset*. Retrieved September 17, 2022, from <https://www.kaggle.com/datasets/elakiricoder/gender-classification-dataset>
- [6] Nawawi, H. M., Purnama, J. J., & Hikmah, A. B. (2019). Komparasi Algoritma Neural Network Dan Naive Bayes Untuk Memprediksi Penyakit Jantung. *Jurnal Pilar Nusa Mandiri*, 15(2), 189–194. <https://doi.org/10.33480/pilar.v15i2.669>
- [7] Suryawan, F. M., Pragantha, J., & Handayani, T. (2022). Pengenalan Karakter Sipi Menggunakan Metode Convolutional Neural Network. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 10(1), 14–18. <https://doi.org/10.24912/jiksi.v10i1.17825>