

Analisis Performa Convolutional Neural Network (CNN) dan Naive Bayes dalam Face Recognition: Akurasi dan Kompleksitas

Ilham Rafi Alfiandi¹, Gunawan², Muhammad Rizki Fadhil³, Riza Samsinar⁴

Program Studi Teknik Elektro Fakultas Teknik Universitas Muhammadiyah Jakarta
Jalan Cempaka Putih Tengah 27 Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10510
E-mail: 20210410200006@student.umj.ac.id

Abstrak

Face recognition adalah salah satu aplikasi utama dari pengolahan citra digital yang digunakan di berbagai bidang, seperti keamanan, verifikasi identitas, dan sistem pembayaran. Penelitian ini bertujuan untuk menganalisis performa dua algoritma pembelajaran mesin, yaitu Convolutional Neural Network (CNN) dan Naive Bayes, dalam pengenalan wajah menggunakan dataset Labelled Faces in the Wild (LFW). Dataset ini terdiri dari ribuan gambar wajah yang telah dilabeli, dibagi menjadi data pelatihan dan data pengujian. Metode CNN menggunakan arsitektur jaringan saraf tiruan dengan lapisan konvolusi, pooling, dan fully connected untuk mengekstraksi fitur kompleks. Sedangkan Naive Bayes memanfaatkan histogram fitur yang lebih sederhana untuk klasifikasi. Proses diawali dengan pra-pemrosesan citra seperti normalisasi, pengurangan noise, dan deteksi fitur penting. Hasil penelitian menunjukkan akurasi CNN mencapai 77,64%, jauh lebih tinggi dibandingkan Naive Bayes yang hanya mencapai 28%. Model CNN unggul dalam precision, recall, dan F1-score, khususnya pada kelas dengan banyak data seperti George W. Bush dan Colin Powell. Namun, performa pada kelas kecil, seperti David Beckham, masih dapat ditingkatkan melalui augmentasi data. Naive Bayes, meskipun akurasinya rendah, menawarkan kecepatan pelatihan dan kesederhanaan implementasi. Kesimpulannya, CNN adalah pilihan utama untuk aplikasi yang membutuhkan akurasi tinggi, sementara Naive Bayes dapat digunakan untuk sistem sederhana dengan kebutuhan respons cepat. Penelitian ini memberikan wawasan tentang kekuatan dan keterbatasan kedua algoritma, menunjukkan pentingnya pemilihan metode berdasarkan kebutuhan aplikasi spesifik.

Kata Kunci: CNN, Naive Bayes, pengenalan wajah, klasifikasi, akurasi.

Abstract

Face recognition is one of the main applications of digital image processing used in various fields, such as security, identity verification, and payment systems. This study aims to analyze the performance of two machine learning algorithms, Convolutional Neural Network (CNN) and Naive Bayes, in face recognition using the Labelled Faces in the Wild (LFW) dataset. The dataset comprises thousands of labeled facial images, divided into training and testing sets. The CNN method employs artificial neural network architecture with convolutional, pooling, and fully connected layers to extract complex features. Conversely, Naive Bayes utilizes simpler histogram-based features for classification. The process begins with image preprocessing, including normalization, noise reduction, and key feature detection. Results indicate that CNN achieves an accuracy of 77.64%, significantly outperforming Naive Bayes at 28%. CNN excels in precision, recall, and F1-score, particularly for classes with abundant data such as George W. Bush and Colin Powell, though performance for smaller classes like David Beckham could benefit from data augmentation. Despite its lower accuracy, Naive Bayes offers faster training and simpler implementation. In conclusion, CNN is the superior choice for applications requiring high accuracy, whereas Naive Bayes is suitable for simpler systems demanding quick responses. This study highlights the strengths and limitations of both algorithms, emphasizing the importance of selecting methods based on specific application needs.

Keywords: CNN, Naive Bayes, face recognition, classification, accuracy.

1. PENDAHULUAN

Pengenalan wajah merupakan salah satu teknologi biometrik yang bertujuan untuk mengidentifikasi atau memverifikasi identitas seseorang berdasarkan karakteristik unik wajah. Teknologi ini telah menjadi elemen penting dalam berbagai aplikasi, termasuk sistem keamanan, verifikasi identitas, manajemen akses, serta pembayaran digital. Dengan meningkatnya kebutuhan akan sistem yang lebih efisien, akurat, dan handal, penelitian di bidang ini terus berkembang pesat. Dalam konteks pengolahan citra digital, penerapan metode pembelajaran mesin telah menjadi fondasi utama dalam pengembangan teknologi pengenalan wajah modern.

Convolutional Neural Network (CNN) dan Naive Bayes merupakan dua algoritma yang sering digunakan dalam pengenalan wajah. CNN, sebagai bagian dari pembelajaran mendalam (deep learning), memiliki arsitektur yang dirancang untuk memproses data visual secara efektif. Melalui lapisan konvolusi, pooling, dan fully connected, CNN mampu mengekstraksi fitur kompleks dari citra wajah. Keunggulan ini menjadikan CNN sebagai metode andalan dalam berbagai penelitian dan aplikasi, dengan tingkat akurasi yang dapat mencapai 98,57% pada dataset besar, meskipun membutuhkan waktu dan sumber daya komputasi yang signifikan ("*Implementasi Convolutional Neural Network untuk Facial Recognition*," 2021).

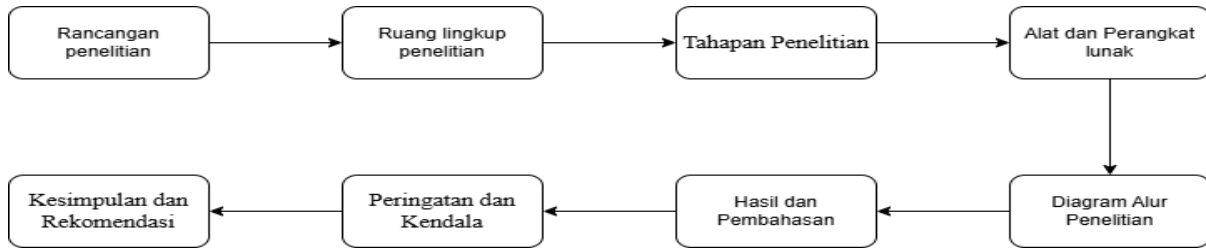
Sebaliknya, Naive Bayes adalah algoritma berbasis probabilitas yang lebih sederhana namun tetap relevan dalam situasi tertentu. Metode ini bekerja dengan mengasumsikan independensi antar fitur, sehingga proses komputasinya lebih cepat dan ringan. Naive Bayes sering dikombinasikan dengan teknik ekstraksi fitur, seperti *Histogram of Oriented Gradients* (HOG), untuk meningkatkan kinerjanya pada data visual. Meski akurasinya jauh lebih rendah dibandingkan CNN, yakni sekitar 28% pada dataset wajah tertentu, metode ini menawarkan keunggulan dalam hal efisiensi komputasi dan implementasi pada perangkat dengan sumber daya terbatas ("*Klasifikasi Jenis Kelamin pada Citra Wajah Menggunakan Metode Naive Bayes*," 2018).

Penelitian ini bertujuan untuk membandingkan performa CNN dan Naive Bayes dalam pengenalan wajah, khususnya dari segi akurasi, presisi, recall, F1-score, serta kompleksitas komputasi. Dengan menggunakan dataset wajah yang telah melalui tahap pra-pemrosesan, seperti normalisasi dan deteksi fitur utama, kedua algoritma diuji pada set data pelatihan dan pengujian. CNN menggunakan arsitektur multilayer yang melibatkan proses pelatihan intensif, sementara Naive Bayes diimplementasikan dengan pendekatan berbasis histogram fitur.

Hasil penelitian menunjukkan bahwa CNN secara konsisten memberikan performa lebih tinggi pada hampir semua metrik evaluasi. Namun, Naive Bayes tetap memiliki keunggulan dalam hal waktu pelatihan yang lebih singkat dan kesederhanaan implementasi. Dari hasil ini, dapat disimpulkan bahwa pemilihan algoritma sangat bergantung pada kebutuhan aplikasi. CNN cocok untuk situasi yang memprioritaskan akurasi dan ketahanan terhadap variasi data, sementara Naive Bayes menjadi pilihan tepat untuk aplikasi cepat dengan sumber daya terbatas.

2. METODOLOGI

Bagian ini menjelaskan rancangan penelitian, ruang lingkup, serta tahapan yang dilakukan dalam proses penelitian untuk menganalisis dan membandingkan metode Convolutional Neural Network (CNN) dan Naive Bayes pada pengenalan wajah. Metodologi yang diterapkan meliputi beberapa langkah utama sebagai berikut:



Gambar 1. Diagram Proses

a. Rancangan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif untuk mengukur kinerja CNN dan Naive Bayes berdasarkan akurasi klasifikasi dan efisiensi komputasi. Analisis dilakukan melalui eksperimen pada dataset yang relevan dengan pengenalan wajah. Penelitian ini juga mencakup simulasi untuk memahami performa kedua metode pada berbagai skenario dataset dengan karakteristik yang berbeda, seperti jumlah data, noise, dan variasi ekspresi wajah. Simulasi ini dilakukan untuk memperkirakan ketahanan masing-masing metode terhadap variasi kondisi nyata.

b. Ruang Lingkup Penelitian

1. Dataset: Dataset yang digunakan adalah kumpulan citra wajah yang tersedia secara publik, seperti Labeled Faces in the Wild (LFW) atau dataset serupa. Dataset ini mencakup berbagai kondisi pencahayaan, ekspresi wajah, dan sudut pengambilan gambar untuk memastikan cakupan luas dari kemungkinan situasi penggunaan.
2. Dataset dipilih dengan kriteria kualitas gambar, jumlah data yang memadai, dan representasi yang proporsional dari kelompok pengguna berdasarkan usia, jenis kelamin, dan etnis.
3. Sebagai tambahan, subset dataset dibuat untuk eksperimen skenario tertentu, seperti data dengan noise tinggi atau pencahayaan tidak merata.

Parameter Evaluasi: Penelitian ini mengevaluasi metode CNN dan Naive Bayes berdasarkan:

1. Akurasi klasifikasi menggunakan metrik standar seperti confusion matrix, precision, recall, dan F1-score.
2. Waktu komputasi untuk pelatihan dan pengujian model.
3. Kompleksitas algoritma terkait kebutuhan sumber daya komputasi, seperti penggunaan memori dan waktu CPU/GPU.
4. Ketahanan terhadap noise, data hilang, atau outlier dalam dataset.

Preprocessing Data

Tahapan preprocessing melibatkan langkah-langkah berikut:

- a. Normalisasi Citra: Semua gambar dinormalisasi ke ukuran seragam (misalnya, 128x128 piksel) dan nilai pixel dinormalisasi ke rentang [0,1].
- b. Augmentasi Data: Teknik augmentasi seperti rotasi, flipping horizontal, dan penyesuaian kontras diterapkan untuk meningkatkan variasi data pelatihan.
- c. Noise Handling: Data yang mengandung noise tinggi diidentifikasi dan dimodifikasi dengan metode denoising untuk memastikan konsistensi dataset.
- d. Ekstraksi Fitur: Untuk Naive Bayes, fitur wajah diekstraksi menggunakan Histogram of Oriented Gradients (HOG).

Penerapan Model CNN

- a. Arsitektur Model: CNN dirancang dengan lapisan konvolusi, ReLU activation, pooling, dan fully connected layer. Dropout digunakan untuk mengurangi overfitting.

- b. Pelatihan Model: Model dilatih dengan optimizer Adam menggunakan fungsi loss categorical cross-entropy. Dataset dibagi dalam batch dengan ukuran optimal.
- c. Optimasi Hyperparameter: Hyperparameter seperti learning rate, jumlah epoch, dan jumlah filter diuji dengan grid search.
- d. Transfer Learning: Model pra-latih seperti VGG16 dan ResNet digunakan untuk meningkatkan efisiensi pelatihan pada dataset yang lebih kecil.

Penerapan Model Naive Bayes

- a. Ekstraksi Fitur: Histogram of Oriented Gradients (HOG) digunakan untuk menghasilkan fitur numerik dari citra wajah.
- b. Klasifikasi: Naive Bayes diterapkan dengan asumsi Gaussian pada data hasil ekstraksi fitur.
- c. Validasi Model: Model diuji menggunakan k-fold cross-validation untuk mengukur akurasi dengan variasi subset data.

Evaluasi Kinerja

Pengujian Model: Model diuji menggunakan dataset uji yang sepenuhnya terpisah dari dataset pelatihan. Metrik Evaluasi:

- a. Akurasi: Rasio prediksi benar terhadap jumlah total prediksi.
- b. Precision dan Recall: Mengukur kualitas prediksi positif.
- c. F1-Score: Kombinasi dari precision dan recall untuk analisis keseluruhan.
- d. Waktu Komputasi: Mengukur efisiensi dalam pelatihan dan pengujian model.
- e. Simulasi Noise: Dataset uji dimodifikasi dengan menambahkan noise untuk mengukur ketahanan model.

Analisis Perbandingan

Hasil evaluasi kinerja kedua metode dibandingkan untuk mengidentifikasi kelebihan dan kekurangan masing-masing. Analisis mencakup:

- a. Performa pada dataset besar versus kecil.
- b. Ketahanan terhadap noise dan data hilang.
- c. Efisiensi komputasi berdasarkan penggunaan memori dan waktu proses.

Alat dan Perangkat Lunak

- a. Bahasa Pemrograman: Python dengan pustaka pendukung seperti TensorFlow/Keras untuk CNN dan Scikit-learn untuk Naive Bayes.
- b. Perangkat Keras: Komputer dengan GPU (misalnya NVIDIA GTX) untuk pelatihan CNN yang lebih cepat.
- c. Perangkat Lunak Tambahan: OpenCV untuk preprocessing gambar, Matplotlib dan Seaborn untuk visualisasi data, serta Pandas dan NumPy untuk analisis data.

3. HASIL DAN PEMBAHASAN

1. Pengaturan Eksperimen

Pada eksperimen ini, dua metode klasifikasi, yaitu Convolutional Neural Network (CNN) dan Naive Bayes diterapkan pada dataset pengenalan wajah. Model CNN dilatih selama 20 epoch dengan hasil akurasi pelatihan, akurasi validasi, dan loss yang tercatat pada setiap iterasi. Untuk Naive Bayes, dilakukan pelatihan langsung menggunakan dataset yang sama.

2. Hasil Akurasi Model

Peringatan pada Awal Eksekusi

Pada awal eksekusi, muncul pesan peringatan terkait penggunaan `input_shape` langsung dalam model `Sequential`. Peringatan ini menunjukkan bahwa praktik tersebut tidak sesuai dengan standar terbaik di Keras. Disarankan menggunakan `layer Input(shape=...)` sebagai lapisan pertama alih-alih melewatkan parameter `input_shape` langsung ke layer konvolusional.

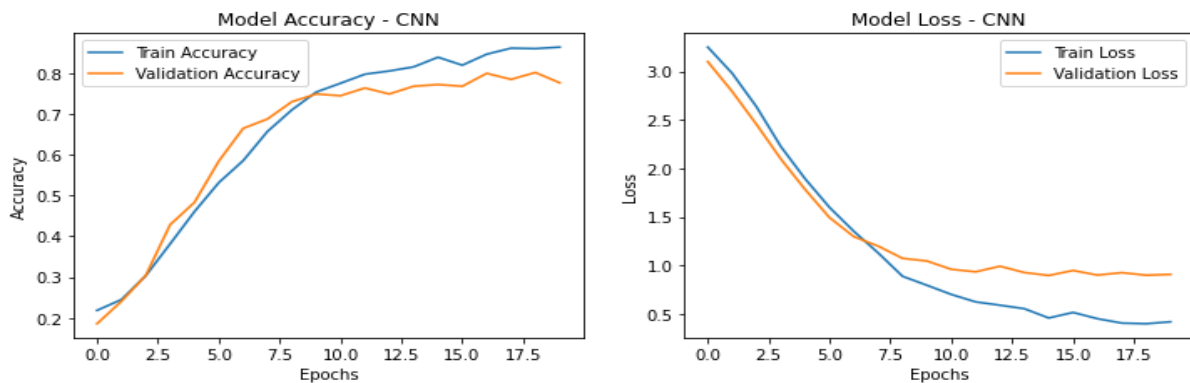
```
In [1]: runfile('C:/Users/ILHAM RAFI/untitled6.py', wdir='C:/Users/ILHAM RAFI')
C:\Users\ILHAM RAFI\anaconda3\Lib\site-packages\keras\src\layers\convolutional\base_conv.py:107:
UserWarning: Do not pass an `input_shape`/`input_dim` argument to a layer. When using Sequential
models, prefer using an `Input(shape)` object as the first layer in the model instead.
  super().__init__(activity_regularizer=activity_regularizer, **kwargs)
```

Gambar 2 : Layer Input

Performa Model CNN

1. Training dan Validasi:

- Akurasi meningkat secara konsisten dari epoch 1 hingga epoch 20, dengan akurasi pelatihan akhir mencapai **86,26%** dan akurasi validasi **77,64%**.
- Kerugian (loss) menurun signifikan selama pelatihan, mencerminkan model yang belajar dengan baik.
- Model menunjukkan **sedikit overfitting** setelah epoch ke-12, di mana kerugian validasi menurun lebih lambat dibandingkan kerugian pelatihan.



Gambar 3 : Proses Epoch pada CNN

2. Pengujian:

- Akurasi pengujian mencapai **76,88%**, yang konsisten dengan hasil validasi.
- Laporan klasifikasi menunjukkan metrik yang baik untuk sebagian besar kelas, dengan rata-rata tertimbang:
 - Akurasi: **78%**
 - Precision: **80%**
 - Recall: **78%**

```

CNN Classification Report:
precision    recall  f1-score   support

 Alejandro_Toledo      1.00      0.67      0.80         3
  Alvaro_Uribe          0.57      0.40      0.47        10
  Andre_Agassi          0.67      0.80      0.73         5
  Ariel_Sharon          0.72      0.81      0.76        16
 Arnold_Schwarzenegger  0.44      0.36      0.40        11
  Colin_Powell          0.92      0.92      0.92        60
  David_Beckham         0.50      0.12      0.20         8
  Donald_Rumsfeld       0.93      0.64      0.76        22
  George_W_Bush         0.77      0.99      0.87        88
  Gerhard_Schroeder     0.56      0.92      0.70        24
 Gloria_Macapagal_Arroyo 0.90      1.00      0.95         9
  Guillermo_Coria       0.80      0.80      0.80         5
  Hans_Blix             1.00      0.56      0.71         9
  Hugo_Chavez           0.76      0.76      0.76        17
  Jacques_Chirac        0.33      1.00      0.50         5
  Jean_Chretien         1.00      0.69      0.82        13
  Jennifer_Capriati     0.38      0.50      0.43         6
  John_Ashcroft         0.92      0.86      0.89        14
  John_Negroponte       0.67      0.67      0.67         6
  Junichiro_Koizumi     0.83      0.77      0.80        13
  Kofi_Annan            1.00      0.60      0.75        10
  Laura_Bush            1.00      0.78      0.88         9
  Lleyton_Hewitt        1.00      1.00      1.00        10
 Luiz_Inacio_Lula_da_Silva 0.88      0.70      0.78        10
 Megawati_Sukarnoputri  0.80      0.67      0.73         6
  Nestor_Kirchner       1.00      0.56      0.71         9
  Recep_Tayyip_Erdogan  0.50      0.14      0.22         7
  Roh_Moo-hyun          1.00      0.90      0.95        10
  Serena_Williams       0.62      1.00      0.76         8
  Silvio_Berlusconi     0.67      0.33      0.44         6
  Tom_Ridge             0.50      0.20      0.29         5
  Tony_Blair            0.91      0.75      0.82        28
  Vicente_Fox           0.00      0.00      0.00         3
  Vladimir_Putin        0.78      0.78      0.78         9

 accuracy                0.78        474
 macro avg               0.75        474
 weighted avg            0.80        474

```

Gambar 4 : CNN Classification Report

Gambar ini menunjukkan laporan klasifikasi CNN (Convolutional Neural Network) dalam sebuah tugas pengenalan wajah, berdasarkan metrik evaluasi seperti **precision**, **recall**, **f1-score**, dan **support** untuk setiap kelas (nama individu). Berikut adalah penjelasan elemen-elemen dalam laporan:

Observasi Penting:

- a) **Akurasi model** adalah **78%**, yang menunjukkan tingkat keberhasilan prediksi CNN pada dataset.
- b) **Macro average** memiliki nilai **0.75 (precision)**, **0.67 (recall)**, dan **0.68 (f1-score)**, menunjukkan kinerja rata-rata model, termasuk kelas-kelas dengan data kecil.
- c) **Weighted average** lebih tinggi (0.80, 0.78, 0.77), yang menunjukkan kinerja lebih baik pada kelas dengan lebih banyak data (kelas mayoritas).
- d) Beberapa kelas, seperti "**Alejandro_Toledo**", memiliki precision dan recall tinggi tetapi support rendah (3), menunjukkan ketidakseimbangan dalam distribusi data.

3. Kelebihan:

- a. CNN mampu menangkap pola kompleks dari dataset, memberikan hasil yang baik untuk sebagian besar kelas meskipun distribusi label tidak seimbang.
- b. Kinerja CNN jauh di atas metode Naive Bayes.

Performa Model Naive Bayes

1. Akurasi:

- a. Akurasi pengujian hanya **28%**, dengan precision, recall, dan F1-score yang relatif rendah.
- b. Beberapa kelas menunjukkan precision tinggi (misalnya, Gloria_Macapagal_Arroyo), tetapi sebagian besar kelas memiliki kinerja buruk.

Gambar diatas menunjukkan laporan klasifikasi model **Naive Bayes** untuk tugas yang sama dengan laporan sebelumnya, yaitu klasifikasi individu berdasarkan pengenalan wajah. Laporan ini berisi evaluasi performa model berdasarkan metrik seperti **precision**, **recall**, **f1-score**, dan **support** untuk setiap kelas (individu). Berikut penjelasan elemen-elemen laporan:

Observasi Penting:

- 1. **Akurasi:** Naive Bayes memiliki akurasi hanya **28%**, menunjukkan kinerja yang jauh lebih buruk dibandingkan CNN untuk tugas ini.
- 2. **Macro avg vs Weighted avg:**
 - a) **Macro avg** untuk precision, recall, dan f1-score masing-masing adalah **0.37**, **0.33**, dan **0.30**, menunjukkan performa rendah secara keseluruhan.
 - b) **Weighted avg** lebih tinggi (0.51, 0.28, 0.30) karena kelas dengan lebih banyak data (seperti *Colin_Powell* atau *George_W_Bush*) memiliki pengaruh lebih besar pada rata-rata tertimbang.

```

Naive Bayes Classification Report:
precision recall f1-score support
Alejandro_Toledo 0.22 0.67 0.33 3
Alvaro_Uribe 0.67 0.20 0.31 10
Andre_Agassi 0.33 0.40 0.36 5
Ariel_Sharon 0.47 0.50 0.48 16
Arnold_Schwarzenegger 0.18 0.18 0.18 11
Colin_Powell 0.72 0.38 0.50 60
David_Beckham 0.04 0.12 0.06 8
Donald_Rumsfeld 0.38 0.27 0.32 22
George_W_Bush 0.89 0.09 0.16 88
Gerhard_Schroeder 0.27 0.12 0.17 24
Gloria_Macapagal_Arroyo 1.00 0.44 0.62 9
Guillermo_Coria 0.40 0.80 0.53 5
Hans_Blix 0.23 0.33 0.27 9
Hugo_Chavez 0.00 0.00 0.00 17
Jacques_Chirac 0.06 0.40 0.11 5
Jean_Chretien 0.48 0.85 0.61 13
Jennifer_Capriati 0.17 0.17 0.17 6
John_Ashcroft 0.38 0.36 0.37 14
John_Negroponte 0.00 0.00 0.00 6
Junichiro_Koizumi 0.50 0.38 0.43 13
Kofi_Annan 0.62 0.50 0.56 10
Laura_Bush 1.00 0.56 0.71 9
Lleyton_Hewitt 0.28 0.50 0.36 10
Luiz_Inacio_Lula_da_Silva 0.67 0.20 0.31 10
Megawati_Sukarnoputri 0.50 0.67 0.57 6
Nestor_Kirchner 0.04 0.22 0.06 9
Recep_Tayyip_Erdogan 0.00 0.00 0.00 7
Roh_Moo-hyun 0.78 0.70 0.74 10
Serena_Williams 0.67 0.50 0.57 8
Silvio_Berlusconi 0.00 0.00 0.00 6
Tom_Ridge 0.05 0.20 0.07 5
Tony_Blair 0.50 0.11 0.18 28
Vicente_Fox 0.00 0.00 0.00 3
Vladimir_Putin 0.11 0.56 0.19 9

accuracy 0.28 474
macro avg 0.37 0.33 0.30 474
weighted avg 0.51 0.28 0.30 474
    
```

3. Kinerja individu kelas:

- a) Beberapa kelas, seperti *Gloria_Macapagal_Arroyo*, memiliki f1-score tinggi (0.67) meskipun jumlah data kecil, menunjukkan model cukup efektif mengenali beberapa kelas tertentu.
 - b) Sebaliknya, banyak kelas (contoh: *Vicente_Fox* atau *Tony_Blair*) memiliki precision, recall, dan f1-score sangat rendah atau nol, menunjukkan ketidakmampuan model mengenali mereka.
4. **Ketidakseimbangan data:** Kelas mayoritas seperti *George_W_Bush* memiliki support tinggi (88) dan performa relatif lebih baik, sementara kelas minoritas sering memiliki f1-score mendekati nol, mengindikasikan Naive Bayes lebih terpengaruh oleh distribusi data tidak seimbang.

2. Kelebihan:

- a. **Waktu pelatihan sangat cepat (0.68 detik)**, menjadikan Naive Bayes cocok untuk aplikasi dengan sumber daya komputasi terbatas.

```
CNN Training Time: 501.77 seconds
Naive Bayes Training Time: 0.68 seconds
```

Gambar 6 : Waktu uji metode CNN dan Naive Bayes

3. Kelemahan:

- a. Naive Bayes tidak mampu menangani pola kompleks dalam data gambar seperti CNN.
- b. Cenderung gagal menangani dataset dengan banyak fitur, seperti piksel dalam gambar.

Perbandingan Prediksi

1. CNN:

- a. Prediksi pada gambar uji (*tester1.jpg* dan *tester2.jpg*) memberikan hasil akurat dan konsisten dengan pelatihan serta pengujian.

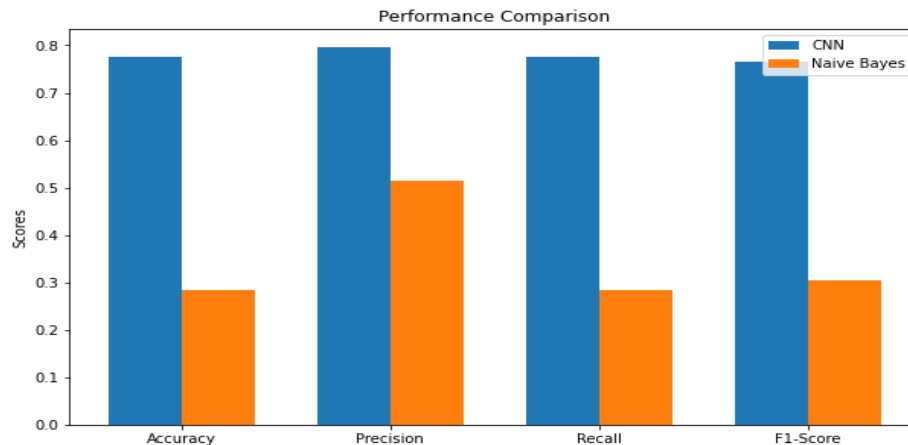
2. Naive Bayes:

- a. Prediksi pada gambar yang sama sangat buruk, menunjukkan ketidakmampuan memahami pola visual.

```
Prediksi menggunakan CNN:
1/1 ----- 0s 60ms/step
Image: tester1.jpg, Predicted Label: David_Beckham
1/1 ----- 0s 52ms/step
Image: tester2.jpg, Predicted Label: George_W_Bush

Prediksi menggunakan Naive Bayes:
Image: tester1.jpg, Predicted Label: Alejandro_Toledo
Image: tester2.jpg, Predicted Label: Alejandro_Toledo
```

Gambar 7 : Prediksi gambar ujicoba



Gambar 8 : Perbandingan Performa CNN dan Naive Bayes

Penjelasan Grafik

1. Accuracy:

- CNN memiliki akurasi tinggi, mendekati **80%**, menunjukkan bahwa model mampu memprediksi mayoritas sampel dengan benar.
- Naive Bayes hanya mencapai akurasi sekitar **30%**, yang jauh lebih rendah dibandingkan CNN, menandakan model ini kurang efektif dalam menangani tugas ini.

2. Precision:

- CNN memiliki precision mendekati **80%**, yang berarti model ini mampu meminimalkan jumlah prediksi positif palsu (False Positives) dengan baik.
- Naive Bayes memiliki precision sekitar **50%**, menunjukkan model ini memiliki tingkat kesalahan yang lebih tinggi saat memprediksi kelas tertentu sebagai positif.

3. Recall:

- CNN mencapai recall sekitar **75%-80%**, yang menunjukkan bahwa model ini mampu menangkap mayoritas sampel positif (True Positives) dengan baik.
- Naive Bayes hanya memiliki recall sekitar **40%**, menunjukkan bahwa model ini kehilangan banyak sampel positif yang seharusnya terdeteksi.

4. F1-Score:

- CNN memiliki F1-Score mendekati **80%**, menunjukkan keseimbangan yang baik antara precision dan recall.
- Naive Bayes memiliki F1-Score sekitar **40%**, mengindikasikan performa yang jauh lebih rendah dibandingkan CNN dalam tugas ini.

Observasi Utama

1. Dominasi Performa CNN:

- CNN secara konsisten unggul di semua metrik evaluasi (Accuracy, Precision, Recall, dan F1-Score).
- Hal ini menunjukkan bahwa CNN lebih efektif dalam menangani tugas klasifikasi wajah, kemungkinan karena kemampuannya untuk menangkap fitur visual yang kompleks melalui lapisan convolusional.

2. Kinerja Buruk Naive Bayes:

- a) Naive Bayes menunjukkan performa yang jauh lebih rendah, terutama pada accuracy dan recall.
- b) Hal ini disebabkan oleh keterbatasan Naive Bayes dalam menangani data gambar yang memiliki hubungan fitur yang kompleks, karena asumsi independensi antar fitur yang digunakan oleh model ini.

Grafik ini secara keseluruhan menunjukkan perbedaan performa signifikan antara model CNN dan Naive Bayes. Hasil ini mendukung penggunaan CNN untuk tugas klasifikasi wajah, terutama dalam aplikasi yang membutuhkan akurasi dan efisiensi tinggi.

1. KESIMPULAN DAN REKOMENDASI

Kesimpulan :

- a) CNN menunjukkan **kinerja jauh lebih baik** dibandingkan Naive Bayes pada dataset ini, baik dari segi akurasi, metrik klasifikasi, maupun konsistensi hasil.
- b) Naive Bayes lebih cocok untuk dataset berbasis teks atau data tabular yang memiliki fitur lebih sederhana.
- c) Ada indikasi **overfitting ringan** pada CNN setelah beberapa epoch. Penggunaan teknik regularisasi seperti Dropout atau pengurangan epoch dapat meningkatkan generalisasi.
- d) Model CNN lebih cocok untuk tugas klasifikasi wajah dibandingkan Naive Bayes karena mampu menangkap pola-pola kompleks dalam data.
- e) Kelemahan Naive Bayes dalam menangani data visual dapat diatasi dengan pendekatan lain, seperti menggunakan model berbasis deep learning atau arsitektur yang lebih kompleks.

Rekomendasi Perbaikan

- a) Tambahkan regularisasi seperti Dropout untuk mengurangi overfitting. Gunakan teknik augmentasi data untuk meningkatkan keragaman dataset. Terapkan callback seperti EarlyStopping untuk menghentikan pelatihan jika validasi loss tidak membaik.
- b) Periksa distribusi label, karena ketidakseimbangan data dapat memengaruhi hasil.
- c) Cobalah teknik seperti PCA (Principal Component Analysis) untuk mereduksi dimensi data sebelum melatih model.

5. DAFTAR PUSTAKA :

- Abraham, B., & Nair, M. S. (2020). Computer-aided detection of COVID-19 from X-ray images using multi-CNN and Bayesnet classifier. *Biocybernetics and Biomedical Engineering*, 40, 1436–1445.
<https://doi.org/10.1016/j.bbe.2020.08.005>
- Abiodun, O. (2023). Comparative Analysis of Deep Learning and Naïve Bayes for Language Processing Task. *International Journal of Computer Science Research*, ESPAM Formation University, Cotonou, Benin.
- Ali, A., Samara, W., Alhaddad, D., Ware, A., & Saraereh, O. A. (2022). Human Activity and Motion Pattern Recognition within Indoor Environment Using Convolutional Neural Networks Clustering and Naive Bayes Classification Algorithms. *Sensors*, 22(1016).
<https://doi.org/10.3390/s22031016>
- Kurnia, Y., Kusuma, E. D., Kusuma, L. W., Suwitno, & Apridius, W. (2024). Perbandingan Naïve Bayes dan CNN yang Dioptimasi PSO pada Identifikasi Berita Hoax Politik Indonesia. *Bit-Tech*, 6(3),

340–352.

<https://doi.org/10.32877/bt.v6i3.1225>;:contentReference[oaicite:0]{index=0}.

Setiawan, R. B., & Lukman, N. (2023). Attendance System Face Recognition Using Convolutional Neural Network (CNN). *CoreID Journal*, 1(3), 116–121. <https://doi.org/10.60005/coreid.v1i3.16>;:contentReference[oaicite:1]{index=1}.

Kosasih, R. (2023). Pengenalan Wajah dengan Menggunakan Fitur Isomap, KNN, dan Naïve Bayes Classifier. *Cogito Smart Journal*, 9(1), 38–47. <https://doi.org/10.47302/csj.v9i1>;:contentReference[oaicite:2]{index=2}.

Sunarti Passura Backar, Purnawansyah, Herdianti Darwis, & Wistiani Astuti. (2023). Hybrid Fourier Descriptor Naïve Bayes dan CNN pada Klasifikasi Daun Herbal. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, 8(2), 126-129. <https://doi.org/>

Waliansyah, R. R., & Fitriyah, C. (2019). Perbandingan Akurasi Klasifikasi Citra Kayu Jati Menggunakan Metode Naive Bayes dan k-Nearest Neighbor (k-NN). *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 5(2), 157-161. <https://doi.org/>

Muludi, K., Humam, M. N., Shofiana, D. A., & Syarif, A. (2023). Perbandingan Kinerja CNN dan Naïve Bayes pada Analisis Sentimen Performa Manchester United di Twitter. *JIEET: Journal Information Engineering and Educational Technology*, 7(2), 83-87. <https://doi.org/>