

Klasifikasi Warna Kulit Pada Sebuah Citra Digital Menggunakan Metode Naive Bayes

Adika Rajendra Haris¹, I Putu Eka Suartana², Bagas Prabaswara Wira N.³, Gede Angga Widya Putra⁴
Program Studi Informatika / Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jakarta
Jl. RS. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12460, Indonesia.
adikarh@upnvj.ac.id, gedeawp@upnvj.ac.id, ipes@upnvj.ac.id, bagaspwn@upnvj.ac.id

Abstrak. Warna Kulit adalah salah satu ciri penting yang membedakan kita semua. Tentunya kita sebagai manusia dapat secara mudah mengklasifikasi dan membedakan warna kulit antar orang tapi sistem digital tidak memiliki mata kita. Kulit penting untuk ekstraksi informasi mengenai kita yang berguna untuk hal seperti *facial recognition* dan *detection*. Komputer perlu dilatih agar dapat melakukan klasifikasi hal yang rumit seperti warna kulit dan wajah. Dari gagasan itu kita bertujuan untuk membuat model klasifikasi warna kulit menggunakan metode klasifikasi Naive Bayes. Kita gunakan kumpulan citra wajah sebagai *dataset* untuk dilatih dan diuji menggunakan Naive Bayes. Python digunakan sebagai bahasa pemrograman yang digunakan untuk melakukan klasifikasi dengan menggunakan berbagai library seperti OpenCV dan scikit learn. Dari hasil percobaan menggunakan Naive Bayes, didapatkan akurasi paling tinggi 75% menggunakan *dataset* yang kita miliki.

Kata Kunci: Warna Kulit, Naive Bayes, Klasifikasi, Citra

1 Pendahuluan

Kulit manusia dapat dijadikan sebagai sumber informasi untuk proses pendeteksian. Informasi yang didapat diambil dari kulit sangat relevan untuk proses pendeteksian antara lain untuk pendeteksian manusia, pendeteksian wajah (*face tracking*), serta penjejakan tangan yang mencakup pengenalan posisi tangan dan gerakannya (*gesture*).Keuntungan yang mampu diperoleh saat menjadikan kulit sebagai sumber awal informasi untuk proses-proses diatas tadi antara lain adalah pemrosesan akan menjadi lebih cepat, tahan (*robust*) terhadap variasi geometris dari pola-pola/tekstur kulit, tahan terhadap perubahan resolusi citra dan mengurangi ketergantungan pada penggunaan peralatan penjejakan yang khusus.

Kulit umumnya dapat dibedakan dengan berapa gelap atau terang kulit tersebut, ini mempermudah proses klasifikasi sejak kita hanya memperhatikan satu aspek atau fitur yaitu warna dari kulit tersebut. Citra kulit kita dapat jadikan citra dan mendapatkan data terang atau gelap piksel tersebut. Dari informasi nilai piksel, kita dapat membedakan warna kulit berdasarkan perbedaan nilai dan jumlah piksel tersebut.

Untuk melakukan proses klasifikasi kita gunakan metode Naive Bayes. Metode ini menggunakan probabilitas dan asumsi bahwa setiap variabel dari *dataset* itu independen dan tidak berkaitan. Dalam kasus ini, kita asumsi bahwa setiap piksel dari 0 hingga 255 itu independen antara satu sama lain. Berdasarkan probabilitas setiap piksel akan digabung untuk mengetahui probabilitas secara keseluruhan. Dari hasil Naive Bayes, kita dapat melakukan perbandingan antara kelas prediksi dari Naive Bayes dan kelas yang kita harapkan. Akurasi diukur dari berapa dari kelas prediksi dan kelas yang diharapkan itu sama atau berbeda.

Banyak penelitian dan riset mengenai klasifikasi warna kulit, salah satunya adalah yang dilakukan oleh Ana Try N, Gede Pasek Sutra, dan L. A. Syamsu Irfan[7][10]. Penelitian yang dilakukan mereka berkait tentang menggunakan metode klasifikasi KNN untuk mengklasifikasikan kulit. Berdasarkan riset tersebut, mereka mendapatkan akurasi sebesar 78%

2 Tinjauan Pustaka

2.1 Melanin Pada Kulit

Di dalam kulit manusia ada melanin yang mempengaruhi warna dari kulit itu sendiri. Melanin merupakan sebuah zat warna alami yang memberi warna pada mata, rambut, dan kulit seseorang. Selain untuk pewarnaan kulit melanin sangat berguna untuk melindungi kulit dari sinar matahari yang dapat menyebabkan kerusakan atau timbulnya penyakit kulit. Sel-sel dalam tubuh yang membentuk melanin disebut dengan melanosit. Melanin pada tiap orang memiliki jumlah yang sama sehingga membuat adanya perbedaan warna kulit, rambut, dan bola mata. Faktor genetik sangat berpengaruh terhadap perbedaan jumlah melanin dalam tubuh. Dengan begitu, seorang anak akan mirip dengan orangtuanya. Dari tiga jenis melanin yang ada pada tubuh manusia hanya jenis Eumelanin memengaruhi sebagian besar warna gelap pada rambut, mata, dan kulit. Terdapat dua jenis eumelanin, yakni eumelanin cokelat dan eumelanin hitam. Perbedaan tersebut berpengaruh terhadap kondisi rambut masing-masing individu[9].

2.2 Citra

Citra merupakan suatu gambaran, kemiripan, atau tiruan dari suatu objek. Citra terbagi menjadi 2 yaitu citra yang memiliki sifat analog dan citra memiliki sifat digital. Citra analog adalah citra yang memiliki sifat *continue* seperti gambar pada monitor televisi, foto sinar X, dan lain-lain. Sedangkan citra digital adalah citra yang bisa diolah oleh komputer. Citra terdiri atas piksel yang mendeskripsi nilai warna dari titik sebuah citra. Citra digital biasanya dapat direpresentasikan dalam bentuk matrix dengan ukuran matriks tersebut merepresentasikan besar dari citra tersebut[2].

2.3 Histogram

Histogram citra merupakan suatu grafik yang merepresentasikan jumlah penyebaran nilai dari sebuah citra. Hal yang biasanya dapat dilihat melewati histogram adalah kontras dari sebuah citra serta frekuensi dari kemunculan piksel. Histogram dapat digunakan sebagai metode pengolahan citra yang kualitatif dan kuantitatif[1].

2.4 Naive Bayes

Algoritma Naive Bayes adalah sebuah metode klasifikasi dengan menggunakan probabilitas dan statistik yang didasari oleh teorema bayes yang ditemukan oleh Thomas Bayes. Untuk memprediksi peluang di masa depan Algoritma Naive Bayes menggunakan pengalaman dari masa sebelumnya. Karakteristik utama Naive Bayes *Classifier* ini adalah kuatnya asumsi (naïf) akan independensi dari masing-masing kondisi / kejadian. Ini berarti peluang dari satu kondisi tidak mengefek ke kondisi lain, tapi dalam aplikasi dunia nyata jarang adanya kasus dimana kondisi itu tidak terkait dengan kondisi lain[6].

2.5 Model Warna

Citra merupakan perhitungan dari intensitas-intensitas cahaya yang disimpan dalam format 8 bit pada spektrum elektromagnetik *single band*. Nilai intensitas piksel dari citra grayscale berdasar pada derajat warna keabuan. terdiri dari berbagai nuansa warna-warna monokromatik dari hitam menjadi putih. Untuk pengolahan citra gambar rentang warna dari hitam dan putih pada citra sangat baik digunakan[8].

3 Metodologi Penelitian

3.1 Identifikasi Masalah

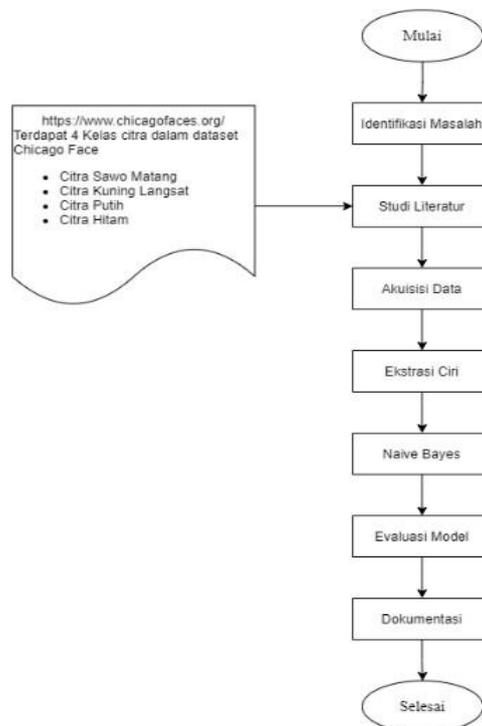
Identifikasi masalah adalah bagian yang berfokus pada penjelasan poin penting yang dibahas dalam penelitian. Yaitu penerapan Algoritma Naive Bayes untuk mengklasifikasi warna kulit manusia.

3.2 Studi Literatur

Studi Literatur umumnya digunakan oleh peneliti untuk dapat mengumpulkan berbagai macam data yang diperlukan dalam sebuah penelitian. Data yang digunakan dalam penelitian ini didapat dari situs web dengan URL <https://www.chicagofaces.org/>.

3.3 Kerangka Berpikir

Kerangka berpikir yang digunakan untuk melakukan penelitian dapat dilihat pada Gambar 1.



Gambar. 1. Kerangka Berpikir

3.4 Akuisisi Data

Data - data wajah yang digunakan untuk klasifikasi citra wajah pada penelitian ini untuk data latih dan data uji diambil dengan proses dokumentasi. Penulis mengambil 4 kelas citra wajah pada situs web dengan URL <https://www.chicagofaces.org/>[3][4][5]. *Dataset* tersebut mengandung hampir 2GB data yang dapat kita gunakan, namun kita sortir 20 yang kita akan gunakan. 4 kelas data yang digunakan dalam penelitian ini, antara lain:

- Citra Kuning Langsung
- Citra Sawo Matang
- Citra Putih
- Citra Hitam

3.5 Ekstraksi Ciri

Ciri yang akan diambil dari *dataset* kita adalah fitur intensitas piksel dari data citra kita. Data intensitas akan kita jadikan histogram, dimana akan dicatat jumlah dari sebuah bit piksel dari 0 hingga 255. Data histogram ini akan dijadikan data yang dilatih dan diuji ke metode klasifikasi Naive Bayes.

3.6 Klasifikasi Naive Bayes

Dari data yang diambil dari ekstraksi ciri kita membagi data tersebut menjadi data latih dan data uji. Data latih akan digunakan untuk melatih model agar dapat memprediksi kelas. Data uji akan digunakan untuk menguji model yang telah dilatih dimana akan menghasilkan prediksi kelas dari data uji tersebut.

3.7 Evaluasi Model

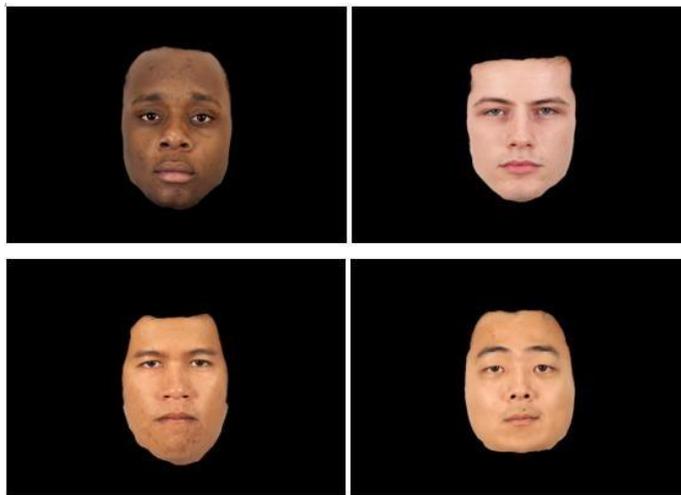
Setelah citra berhasil diklasifikasikan dengan Naive Bayes, selanjutnya akan diukur hasil keluarannya untuk menentukan tingkat akurasi dari pengujian naive bayes. Ini dilakukan dengan membandingkan data kelas prediksi dengan kelas yang seharusnya. Akurasi diambil dari persentase yang sama dan berbeda.

3.8 Dokumentasi

Pada tahap ini dapat ditarik kesimpulan akhir dari hasil pengujian algoritma Naive Bayes terhadap warna kulit manusia. Setelah semua tahapan diselesaikan maka tahap yang terakhir adalah melakukan dokumentasi pada seluruh kegiatan yang kita lakukan, lalu ditulis menjadi laporan yang disusun secara sistematis.

4 Hasil dan Pembahasan

4.1 Hasil



Gambar. 2. *Dataset* 4 jenis ras warna kulit

Untuk data yang digunakan untuk latih dan uji, kita menggunakan total 80 data dengan 20 data untuk setiap kelas. Terdapat 4 kelas yang digunakan untuk model ini. Berikut adalah contoh *dataset* yang kita gunakan dari empat kelas tersebut. *Dataset* kita telah pra-proses dengan mengisolasi wajah mereka agar mendapatkan hasil akurasi terbaik. Kita mencoba untuk melakukan klasifikasi data menggunakan dua tipe citra, yaitu *grayscale* dan RGB. Berdasarkan klasifikasi dari dua model warna citra, kita mendapatkan hasil sebagai berikut.

```
Data test      : [3. 2. 2. 4. 3. 4. 4. 4. 3. 4. 3. 4. 1. 4. 3. 2. 1. 2.]
Data Predict NB : [2. 4. 2. 4. 3. 4. 4. 4. 3. 4. 3. 4. 1. 4. 3. 2. 1. 2.]
Akurasi GNB = 90.0%
Score Cross Val Average: 0.775
Data Predict MGNB: [3. 2. 2. 4. 3. 4. 4. 4. 3. 4. 2. 4. 1. 4. 3. 2. 1. 3. 1. 3.]
Akurasi MGNB= 85.0%
Score Cross Val Average: 0.775
```

Gambar. 3. Hasil Klasifikasi RGB

Gambar diatas adalah hasil dari klasifikasi menggunakan Gaussian Naïve Bayes dan Multinomial Naïve Bayes dengan *dataset* citra RGB.

```
[3. 2. 2. 4. 3. 4. 4. 4. 3. 4. 3. 4. 1. 4. 3. 2.]
[2. 2. 2. 4. 4. 4. 4. 4. 2. 3. 2. 4. 4. 4. 3. 2.]
Score Cross Val Average: 0.625
Akurasi GNB = 62.5%
```

Gambar. 4. Hasil Klasifikasi Gaussian Naïve Bayes

```
[3. 2. 2. 4. 3. 4. 4. 4. 3. 4. 3. 4. 1. 4. 3. 2.]  
[3. 2. 2. 4. 4. 3. 4. 4. 3. 4. 2. 4. 1. 4. 4. 2.]  
Score Cross Val Average: 0.65  
Akurasi MNB = 75.0%
```

Gambar. 5. Hasil Klasifikasi Multinomial Naïve Bayes

Gambar diatas merupakan hasil dari klasifikasi yang menggunakan citra *grayscale*, serta akurasinya. Metode yang digunakan adalah Gaussian Naïve Bayes dan Multinomial Naïve Bayes

4.2 Pembahasan

4.2.1 RGB

Terdapat 20 data yang akan diuji coba dengan dua tipe citra yakni RGB dan *grayscale*. Hasil pertama itu untuk yang RGB. Pada hasil Gaussian Bayes, akurasi dari hasil prediksi yang didaapat sebesar 90%, dengan nilai *Cross Validation* sebesar 0.775 atau 77.5%. Sedangkan, jika menggunakan Multinomial Bayes menghasilkan akurasi yang lebih kecil 5% yaitu sebesar 85%, dengan nilai *Cross Validation* yang sama dengan Gaussian Bayes yaitu sebesar 0.775 atau 77.5%.

4.2.2 Grayscale

Grayscale juga diklasifikasi menggunakan Gaussian Naïve Bayes dan Multinomial Naïve Bayes. Jika menggunakan Gaussian bayes. Akurasi dari hasil prediksi dan hasil seharusnya adalah 62.5%. Menggunakan *Cross Validation* dengan 5 iterasi memiliki rata-rata yang sama. Menggunakan Multinomial Bayes akan menghasilkan nilai akurasi yang lebih tinggi dibanding dengan Gaussian dengan nilai akurasi 75% dan Cross Val rata-rata 0.65.

4.2.3 Analisa

Bedasarkan hasil dari kedua RGB dan *grayscale* dapat dikatakan bahwa menggunakan citra RGB menghasilkan akurasi yang lebih baik dibanding dengan *grayscale*. RGB menggunakan Gaussian dan Multinomial mendapatkan akurasi diatas 80% serta *Cross Validation* 77.5%, lebih besar dari *grayscale* yang akurasi dan *Cross Validation* bernilai sekitar 60 hingga 75% dan 60% untuk *Cross Validation*. Alasan ini adalah karena menggunakan RGB kita dapat membedakan antara warna kulit secara lebih efektif dengan adanya tiga *channel* warna. *Grayscale* hanya menghitung keabuan dari sebuah citra maka dapat dibilang kurang efektif dalam membedakan warna kulit.

5 Kesimpulan Dan Saran

5.1 Kesimpulan

Bedasarkan dari hasil yang kita telah dapat, dapat disimpulkan bahwa klasifikasi warna kulit menggunakan Naïve Bayes dapat dilakukan secara efektif. Tentu ini jika data yang kita gunakan itu berkualitas bagus dan setara. Selain dari itu dapat disimpulkan bahwa menggunakan citra RGB digolong lebih efektif dalam mendeteksi perbedaan warna kulit karena variasi warna piksel dapat membantu akurasi dalam membedakan warna kulit.

5.2 Saran

Saran mengenai riset ini sebagai berikut:

1. Menggunakan data yang perbedaanya lebih terlihat
2. Memakai lebih banyak tipe klasifikasi untuk mendapatkan metode yang lebih optimum
3. Membuat data yang digunakan untuk *training* dan *testing* lebih rata dan adil

Referensi

- [1] Darma Putra. (2010). *Pengolahan Citra Digital*, Westriningsih, Ed. Yogyakarta: Andi.
- [2] Sutojo, Siswanto. (2004). *Membangun Citra Perusahaan*. Jakarta: Damar Mulia Pustaka.
- [3] Ma, Correll, & Wittenbrink (2015). The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data. *Behavior Research Methods*, 47, 1122-1135.
- [4] Lakshmi, Wittenbrink, Correll, & Ma (2021). The India Face Set: International and Cultural Boundaries Impact Face Impressions and Perceptions of Category Membership. *Frontiers in Psychology*, 12, 161.
- [5] Ma, Kantner, & Wittenbrink (2020). Chicago Face Database: Multiracial Expansion. *Behavior Research Methods*.
- [6] Webb, Geoffrey I., Eamonn Keogh, & Risto Miikkulainen (2010). "Naïve Bayes." *Encyclopedia of machine learning* 15, 713-714.
- [7] Ana Try N., I Gede Pasek Suta W., & L. A. Syamsu Irfan (2017). KLASIFIKASI WARNA KULIT MANUSIA DENGAN ALGORITMA K-NEAREST NEIGHBOUR (KNN). *Dielektrika*, 4, 135-141
- [8] Nikolskaia, K., Ezhova, N., Sinkov, A., & Medvedev, M. (2018). Skin detection technique based on HSV color model and SLIC segmentation method. In *Proceedings of the 4th Ural Workshop on Parallel, Distributed, and Cloud Computing for Young Scientists, Ural-PDC 2018. CEUR Workshop Proceedings*, 2281, 123-135.
- [9] Jablonski, Nina G (2004). The evolution of human skin and skin color. *Annual review of Anthropology*, 585-623.
- [10] Dino, Hivi Ismat, & Maiwan Bahjat Abdulrazzaq (2019). Facial expression classification based on SVM, KNN and MLP classifiers. *2019 International Conference on Advanced Science and Engineering (ICOASE). IEEE*, 2019.