

**Pengenalan Wajah Menggunakan Metode
Linear Discriminant Analysis dan C4.5**

Oleh

ZAINAL KARIM

T3116024

SKRIPSI

Untuk memenuhi salah satu syarat ujian
guna memperoleh gelar Sarjana



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2023**

PERSETUJUAN SKRIPSI

Pengenalan Wajah Menggunakan Metode Linear Discriminant Analysis dan C4.5

Oleh

ZAINAL KARIM

T3116024

SKRIPSI

Untuk memenuhi salah satu syarat ujian
guna memperoleh gelar Sarjana
dan telah disetujui oleh Tim pembimbing pada tanggal
Gorontalo, 29 Mei 2023

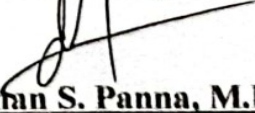
Pembimbing I



Irma Surya Kumala Idris, M.Kom

NIDN : 0921128801

Pembimbing II



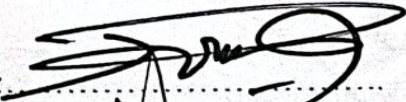
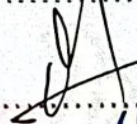
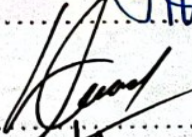
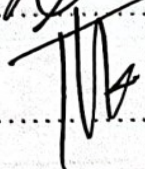
Sudirman S. Panna, M.Kom

NIDN : 0924038205

Pengenalan Wajah Menggunakan Metode Linear Discriminant Analysis dan C4.5

Oleh
ZAINAL KARIM
T3116024

Diperiksa oleh Panitia Ujian Strata satu (S1)
Universitas Icsan Gorontalo
Gorontalo, 29 Mei 2023

1. Ketua Penguji
Irvan Abraham Salihi, M.Kom : 
2. Anggota
Sudirman S.Panna, M.Kom : 
3. Anggota
Irma Surya Kumala Idris, M.Kom: 
4. Anggota
Husdi, M.Kom : 
5. Anggota
M. Effendi Lasulika, M.Kom : 

Mengetahui :

Dekan Fakultas Ilmu Komputer


Irvan Abraham Salihi, M.Kom
NIDN.0928028101

Ketua Program Studi


Sudirman S. Panna, M.Kom
NIDN : 0924038205

PERNYATAAN SKRIPSI

Dengan ini saya menyatakan bahwa :

1. Karya tulis (Skripsi) saya ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) baik di Universitas Ichsan Gorontalo maupun di perguruan tinggi lainnya.
2. Karya tulis (Skripsi) saya ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan dari Tim Pembimbing.
3. Karya tulis (Skripsi) saya ini tidak terdapat karya atau pendapat yang telah dipublikasi orang lain, kecuali secara tertulis dicantumkan sebagai acuan/sitasi dalam naskah dan dicantumkan pula dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi akademik berupa pencabutan gelar, yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma - norma yang berlaku di Universitas Ichsan Gorontalo.

Gorontalo, Mei 2023

Yang Membuat Pernyataan,



Zainal Karim

ABSTRACT

ZAINAL KARIM. T3116024. FACIAL RECOGNITION USING LINEAR DISCRIMINANT ANALYSIS AND C4.5 METHOD

Human face recognition is a field that is developing in the current era. The various applications of tools with that kind of capability have broad usefulness in searching for criminals, criminalization, financial access systems, and human-computer interaction. This application uses a camera on a smartphone to capture a person's face and then compare it with faces that have previously been stored and trained in a database to make the face identification successful. If it does not match, it will declare a failure. The expression of facial recognition is carried out using the Linear Discriminant Analysis method as facial feature extraction to obtain parameter values to be used as input for classification. The facial recognition algorithm used is the C4.5 algorithm because it depends on lighting, the distance between subject and camera, facial angle of view, facial expressions, accessories uses, and background changes.

Keywords: face recognition, Linear Discriminant Analysis, C4.5 algorithm

ABSTRAK

ZAINAL KARIM. T3116024. PENGENALAN WAJAH MENGUNAKAN METODE LINEAR DISCRIMINANT ANALISYST DAN C4.5

Pengenalan wajah manusia (face recognition) merupakan salah satu bidang yang cukup berkembang di era sekarang ini. Berbagai aplikasi dari alat dengan kemampuan seperti ini terbentang luas dari pencarian penjahat, kriminalisasi, sistem akses keuangan, sampai interaksi manusia dengan komputer. Aplikasi ini menggunakan sebuah kamera yang ada pada smartphone untuk menangkap wajah seseorang kemudian dibandingkan dengan wajah yang sebelumnya telah disimpan dan dilatih dalam database. Maka identifikasi wajah berhasil, jika tidak cocok maka akan dinyatakan gagal. Pengenalan ekspresi wajah dilakukan dengan menggunakan metode *Linear Discriminant Analysis* sebagai fitur ekstraksi pada wajah untuk mendapatkan nilai-nilai parameter yang akan dijadikan inputan pada klasifikasi. Algoritma pengenalan wajah yang digunakan adalah Algoritma C4.5 karena bergantung pada pencahayaan, jarak antara subjek dan kamera, sudut pandang wajah, ekspresi/mimik wajah, pemakaian aksesoris, dan perubahan latar (*background*).

Kata kunci: pengenalan wajah, *Linear Discriminant Analysis*, algoritma C4.5

KATA PENGANTAR

Alhamdulillah, penulis dapat menyelesaikan Skripsi ini dengan judul **“Pengenal wajah menggunakan metode *Linear Discriminant Analysis* dan C4.5”** untuk memenuhi salah satu syarat penyusunan Skripsi Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.

Penulis menyadari sepenuhnya bahwa Skripsi ini tidak mungkin terwujud tanpa bantuan dan dorongan dari berbagai pihak, baik bantuan moril maupun materil. Untuk itu, dengan segala keikhlasan dan kerendahan hati, penulis mengucapkan banyak terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Dr. Drs. Juriko Abdussamad, M.Si, selaku Ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Dr. Abdul Gaffar La Tjokke, M.Si, selaku Rektor Universitas Ichsan Gorontalo;
3. Irvan Abraham Salihi, S.Kom, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
4. Sudirman Melangi, S.Kom, M.Kom, selaku Wakil Dekan I Bidang Akademik dan Bidang Kemahasiswaan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
5. Irma Surya Kumala Idris, S.Kom, M.Kom, selaku Wakil Dekan II Bidang Administrasi Umum dan Keuangan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
6. Sudirman S. Panna, S.kom, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
7. Husdi S.Kom, M.Kom, selaku Sekertaris Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;

8. Irma Surya Kumala Idris, S.Kom, M.Kom, selaku Pembimbing I yang telah membimbing penulis selama mengerjakan Skripsi ini;
9. Sudirman S. Panna, S.Kom, M.Kom, selaku Pembimbing II yang telah membimbing penulis selama mengerjakan Skripsi ini;
10. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang telah mendidik dan mengajarkan berbagai disiplin ilmu kepada penulis;
11. Kepada Kedua Orang Tua saya tercinta, atas segala kasih sayang, jerih payah dan do'a restunya dalam membesarkan dan mendidik penulis;
12. Rekan-rekan seperjuangan yang telah banyak memberikan bantuan dan dukungan moril yang sangat besar kepada penulis;
13. Kepada semua pihak yang ikut membantu dalam penyelesaian Skripsi ini yang tak sempat penulis sebutkan satu-persatu.

Semoga Allah, SWT melimpahkan balasan atas jasa-jasa mereka kepada kami. Penulis menyadari sepenuhnya bahwa apa yang telah dicapai ini masih jauh dari kesempurnaan dan masih banyak terdapat kekurangan. Oleh karena itu, penulis sangat mengharapkan adanya kritik dan saran yang konstruktif. Akhirnya, penulis berharap semoga hasil yang telah dicapai ini dapat bermanfaat bagi kita semua, Aamin;

Gorontalo, Desember 2023

Penulis

DAFTAR ISI

PERSETUJUAN SKRIPSI.....	ii
PERNYATAAN SKRIPSI.....	iv
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah	3
1.3 Rumusan Masalah	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
BAB II LANDASAN TEORI	4
2.1. Tinjauan studi	4
2.2. Tinjauan Pustaka	5
2.2.1. Pengenalan Wajah	5
2.2.2. Pengolahan Citra	7
2.2.3. Linear Discriminant Analyst (LDA)	8
2.2.4. Algoritma C4.5	11
2.2.5. Penerapan Algoritma C4.5 Memprediksi Penjualan Wallpaper	12
2.2.6 Evaluasi Model.....	15
2.3. Kerangka Pikir.....	17
BAB III METODE PENELITIAN	18
3.1 Jenis, Metode, Subjek, dan Objek	18
3.1.1. Jenis	18
3.1.2. Metode.....	18
3.1.3. Subjek.....	18

3.1.4. Objek	18
3.2 Pengumpulan Data	18
3.3 Pemodelan	19
3.3.1. Pra Pengolahan Data.....	20
3.3.2. Ekstraksi fitur Linear Discriminant Analisis (LDA)	20
3.3.3. Klasifikasi C4.5	20
3.3.4. Evaluasi Model.....	20
BAB IV HASIL PENELITIAN.....	21
4.1. Hasil Pengumpulan Data.....	21
4.2 Analisis dan Pemodelan	22
4.2.1. Pra Pengolahan Data.....	22
4.2.2 Ekstraksi Linear Discriminant Analisis (LDA).....	25
4.2.3 Klasifikasi C4.5	34
4.3. Confusion Matrix.....	40
BAB V PEMBAHASAN	42
5.1 Objek penelitian.....	42
5.2 Pembahasan Model.....	42
BAB VI PENUTUP	44
6.1. Kesimpulan.....	44
6.2. Saran.....	44
DAFTAR PUSTAKA	45
KODE PROGRAM.....	Error! Bookmark not defined.
DATA.....	51

DAFTAR GAMBAR

Gambar 2.1 Gambar 2.1 Arah optimal vector w sehingga jarak antar mean maksimal dan lebar varian	9
Gambar 2.2 Hasil ekstraksi fitur citra data base	11
Gambar 2.3 Decision Tree	15
Gambar 2.4 Decision Tree dengan Rapid Miner	15
Gambar 2.5 Bagan Kerangka Pikir	17
Gambar 2.6 Pemodelan	19
Gambar 4.1 Hasil Pengumpulan Dataset	21
Gambar 4.1.2 Filter 3 x 3	23
Gambar 4.1.3 Pixel Citra	23
Gambar 4.1.4 kernel/filter yang digunakan	24
Gambar 4.1.5 matriks hasil proses morfologi	24
Gambar 4.1.6 hasil akhir proses morfologi	24
Gambar 4.2.1 hasil normalisasi citra	25
Gambar 2.6 Pemodelan	19
Gambar 2.6 Pemodelan	19

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	4
Tabel 2.2 Data Walpaper.....	112
Tabel 2.3 Nilai Gain dan Nilai Entropy.....	14
Tabel 4.1 Hasil Ekstraksi Fitur	125
Tabel 4.2 Hasil Perhitungan pada <i>Dataset</i>	126
Tabel 4.3 Analisis Atribut, Nilai, Banyaknya Kejadian Nilai, Entropy dan Gain	127
Tabel 4.4 Hasil eksekusi nilai gain tertinggi	128
Tabel 4.5 Hasil akhir nilai gain terendah.....	128
Tabel 4.6 Hasil perhitungan C4.5 untuk target dan prediksi.....	129

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengenalan wajah merupakan suatu system biometrik yang banyak digunakan untuk identifikasi personal pada pengawasan atau akses masuk dalam suatu system keamanan yang bersifat rahasia. Biometrik adalah ciri khusus yang ada pada manusia seperti sidik jari, retina mata, gigi, dan wajah. Pengenalan wajah digunakan untuk mencocokkan data wajah pada citra dengan data wajah yang sudah ada pada *database*, sehingga identitas pemilik wajah dapat diketahui oleh system.

Namun sebelum data wajah diidentifikasi, ada beberapa proses yang harus di lalui salah satunya yaitu proses ekstraksi fitur. Ekstraksi fitur merupakan tahapan mengekstrak ciri atau informasi dari objek didalam citra yang ingin di kenali atau dibedakan dengan objek lainnya.

Ada banyak metode yang digunakan untuk ekstraksi fitur pengenalan wajah salah satunya *Principal Component Analysis* (PCA). Metode PCA bertujuan untuk mereduksi dimensi dengan melakukan transformasi linear dari suatu ruang yang berdimensi tinggi ke dalam ruang berdimensi rendah. Namun kelemahan dari metode PCA adalah kurang optimal dalam pemisahan antar kelas sehingga dapat mempengaruhi tingkat akurasi pengenalan wajah[1].

Untuk mengatasi masalah tersebut, Turk dan Pentland pada tahun 1991 [2] memperkenalkan *Linear Discriminant Analysis* (LDA) untuk pengenalan wajah. Metode ini bekerja dengan menemukan subruang linear yang memaksimalkan perpisahan dua kelas pola menurut Fisher Criterion (bobot fisher kriteria). Hal ini dapat diperoleh dengan meminimalkan jarak matriks sebaran within-class S_w dan memaksimalkan jarak matriks sebaran beetwen-class S_b secara simultan sehingga menghasilkan fisher criterion yang maksimal. *Linear Discriminant Analysis* (LDA) akan menemukan subruang linear dimana kelas-kelas saling terpisah dengan memaksimalkan fisher criterion. Metode *Linear Discriminant Analysis* (LDA) digunakan dalam penelitian ini sebagai metode ekstraksi fitur untuk

mengoptimalkan pemisahan antar kelas citra wajah sehingga hasil pengenalannya lebih akurat.

Setelah melalui tahapan ekstraksi fitur, fitur-fitur penting dari citra wajah nantinya akan digunakan untuk proses klasifikasi. Metode klasifikasi yang akan digunakan dalam penelitian ini adalah C4.5. Metode C4.5 yaitu sebuah algoritma yang digunakan untuk membangun *decision tree* (pengambilan keputusan). Algoritma C4.5 adalah salah satu algoritma induksi pohon keputusan yaitu ID3 (Iterative Dichotomiser 3). ID3 dikembangkan oleh J.Ross Quinlan. Dalam prosedur algoritma ID3, input berupa *sample training*, *label training* dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3. Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi *missing value*, bisa mengatasi *continuu data*, dan *pruning* [3].

Pada penelitian sebelumnya yang dilakukan oleh Fandiansyah dkk dengan judul “Pengenalan wajah menggunakan metode linear discriminant analysis dan k nearest neighbor” berdasarkan hasil pengujian diperoleh akurasi pengenalan citra wajah normal mencapai 98.33% dan akurasi pengenalan citra wajah yang diberi noise sebesar 86,66% [4]. Penelitian juga dilakukan oleh Hizas Sabila Rasyad, Febryanti Sthevanie dan Anditya Arifanto dengan judul “Pengenalan wajah menggunakan metode local binary pattern dan principal component analysis”. Penelitian tersebut menggunakan resolusi rendah telah mencapai tingkat akurasi 94%.

Seluruh citra wajah yang digunakan sebagai data latih dan data uji diambil dari data public. Data yang digunakan sebanyak 45 citra wajah dengan format *.jpg dan berasal dari 15 citra wajah manusia. Setiap individu masing masing mempunyai 3 citra wajah dengan ekspresi yang berbeda.

Penelitian ini menggunakan *Linear Discriminant Analysis* sebagai metode ekstraksi fitur yang diawali dengan memasukkan data wajah kemudian di ubah ke *eigenface* dan selanjutnya di ekstraksi dengan menggunakan *Linear Discriminant Analysis* (LDA). Sedangkan C4.5 akan digunakan untuk klasifikasi fitur-fitur yang telah di ekstraksi, karena metode C4.5 merupakan salah satu metode komputasi yang cukup akurat dalam menyelesaikan masalah klasifikasi.

Berdasarkan latar belakang diatas, maka peneliti melakukan penelitian dengan judul: “**Pengenalan Wajah Menggunakan *Linear Discriminant Analysis* dan *C4.5***”. Diharapkan penelitian ini mampu memberikan hasil akurat untuk fitur pengenalan wajah.

1.2 Identifikasi Masalah

Berdasarkan latar belakang diatas dapat diidentifikasi masalah yang terkait dengan penelitian ini antara lain.

1. Kurangnya tingkat akurasi metode *Principal Component Analysis* (PCA) pada pengenalan wajah.
2. Metode *C4.5* belum pernah digunakan dalam pengenalan wajah.

1.3 Rumusan Masalah

1. Bagaimana kinerja *Linear Discriminant Analysis* dalam mengekstrak fitur statistik pada wajah ?
2. Bagaimana tingkat akurasi metode *C4.5* untuk pengenalan wajah ?

1.4 Tujuan Penelitian

1. Untuk mengetahui kinerja *Linear Discriminant Analysis* dalam mengekstrak fitur pada wajah.
2. Untuk mengetahui tingkat akurasi metode *C4.5* pada pengenalan wajah.

1.5 Manfaat Penelitian

1. Manfaat Teoritis

Teknologi, khususnya dalam bidang *computer vision*, yaitu berupa uji coba metode untuk proses pengenalan wajah.

2. Manfaat Praktis

Sumbangan Pemikiran , Karya , Bahan pertimbangan , atau solusi bagi *software developer* guna mendukung pengambilan keputusan dalam rangka menghasilkan perangkat lunak yang dapat membantu dalam mengenali wajah seseorang.

BAB II

LANDASAN TEORI

2.1. Tinjauan studi

Tabel dibawah ini merupakan penelitian terkait yang berhubungan dengan metode pengenalan wajah:

Tabel 2.1 Penelitian Terkait

NO	PENELITI	JUDUL	TAHUN	METODE	HASIL
1.	Perani Rosyani	Pengenalan wajah menggunakan metode <i>Principal Component Analysis</i> dan <i>Canberra Distance</i>	2017	<i>Principal Component Analysis</i> dan <i>Canberra Distance</i>	Hasil dari penelitian ini adalah tingkat kesamaan yang dihasilkan setelah proses reduksi dan ekstraksi menggunakan PCA didapatkan rata-rata nilai untuk <i>Canberra Distance</i> adalah 77,59.
2.	Hizas Febryanti Anditya	Pengenalan wajah menggunakan metode <i>Local Binary Pattern</i> dan <i>Principal Component Analysis</i> .	2018	<i>Local Binary Pattern</i> (LBP) dan <i>Principal Component</i>	Pengenalan wajah menggunakan metode <i>Local Binary Pattern</i> dan dikombinasikan dengan <i>Principal Component</i>

				<i>Analysis (PCA)</i>	<i>Analysis</i> mencapai akurasi 94% pada resolusi rendah.
3.	Fandiyansyah Jayanti Ika	Pengenalan wajah menggunakan metode <i>Linear Discriminant Analysis</i> dan <i>K Nearest Neighbor</i> .	2017	<i>Linear Discriminant Analysis</i> dan <i>K Nearest Neighbor</i>	Hasil pengujian menunjukkan metode <i>Linear Discriminant Analysis</i> dan <i>K Nearest Neighbor</i> cukup optimal untuk melakukan pengenalan wajah dengan akurasi pengenalan citra wajah normal mencapai 98.33% dan akurasi pengenalan citra wajah yang diberi noise sebesar 86,66%.

2.2. Tinjauan Pustaka

2.2.1. Pengenalan Wajah

Pengenalan wajah merupakan sebuah system identifikasi pribadi yang menggunakan karakteristik wajah seseorang. Pengenalan wajah sendiri merupakan suatu cabang ilmu biometrik, yaitu suatu bidang keilmuan yang menggunakan karakteristik fisik dari seseorang untuk menentukan atau mengungkapkan identitasnya [5]. Secara umum, ada 4 tahapan dalam system pengenalan wajah yaitu akuisisi citra wajah, preprocessing, ekstraksi fitur dan klasifikasi citra wajah

tersebut. Dari keempat tahapan tersebut, tahapan ekstraksi fitur dan klasifikasi merupakan tahapan yang paling penting dalam system pengenalan wajah [4].

1. Akuisisi citra adalah tahapan awal untuk mendapatkan citra digital. Tujuan akuisisi citra adalah menentukan data yang diperlukan dan memilih metode perekaman citra digital. Tahap ini dimulai dari objek yang akan diambil gambarnya, persiapan alat-alat, sampai pada pencitraan. Pencitraan adalah kegiatan transformasi dari citra tampak (foto, gambar, lukisan, patung, pemandangan, dan lain-lain) menjadi citra digital) [6].
2. *Pre-processing* merupakan proses yang dilakukan untuk menghasilkan ciri atau fitur penting dari sebuah citra. Proses ini meliputi peningkatan kualitas citra (kontraks, kecerahan, dan lain-lain), penghilangan noise, perbaikan citra (image restoration), transformasi (image transformation) dan penentuan bagian citra yang akan diobservasi. Pre-processing berfungsi untuk menyeragamkan ukuran citra wajah dan memperbaiki kualitasnya sebelum melalui tahap ekstraksi fitur.
3. Ekstraksi fitur adalah suatu proses pengambilan ciri atau fitur dari citra yang nilainya akan digunakan untuk proses klasifikasi (pengenalan). Ekstraksi fitur dapat dibedakan menjadi 3 jenis (Kalel, Pisal & Bagawade, 2016) :

- a. Ekstraksi Warna

Ruang warna mewakili warna dalam bentuk nilai intensitas. Kita bisa menentukan memvisualisasikan dan menciptakan warna dengan menggunakan metode ruang warna. Metode yang sering digunakan untuk ekstraksi warna antara lain *Histogram Intersection*, *Zernike Chromaticity*, *Distribution Moments* dan *Color Histogram*.

- b. Ekstraksi Bentuk

Bentuk adalah sumber informasi utama yang bisa digunakan untuk pengenalan suatu abjek. Tanpa bentuk, objek konten visual tidak bisa dikenali dengan baik. Dua objek tidak bisa memiliki bentuk yang sama persis tapi dengan menggunakan berbagai algoritma kita bisa mengenali bentuk yang serupa dengan mudah. Metode yang sering digunakan untuk ekstraksi bentuk antara lain Binary Image dan segmentasi secara horizontal dan vertikal.

c. Ekstraksi Tekstur

Tekstur mengandung informasi penting tentang susunan dasar permukaan objek. Metode yang sering digunakan untuk ekstraksi tekstur antara lain *Gray Level Co-Occurrence Matrix* (GLCM), *Gray Level Run Length Matrix* (GLRLM), *Linear Discriminant Analysis* (LDA) dan *Edge Detection*.

4. Klasifikasi citra wajah (pengenalan) merupakan teknik yang digunakan untuk menentukan item dari suatu dataset kedalam kategori atau kelas tertentu. Tujuan klasifikasi adalah memprediksi kelas target secara akurat disetiap kasus yang terdapat dalam data. Pada tahap ini, pola data wajah yang dihasilkan pada tahap pengenalan dibandingkan dengan data pola wajah yang tersimpan pada basis data wajah untuk menemukan pola wajah yang dikenali.

2.2.2. Pengolahan Citra

Seringkali citra mengalami penurunan mutu (degradasi), misalnya mengandung cacat (*noise*), warnanya terlalu kontras, kurang tajam, kabur, dan sebagainya. Citra seperti ini lebih sulit untuk diinterpretasi karena informasi yang disampaikan oleh citra tersebut berkurang. Agar lebih mudah diinterpretasi, maka citra tersebut perlu dimanipulasi menjadi citra baru yang memiliki kualitas lebih baik. Bidang studi yang menyangkut hal ini adalah pengolahan citra [7].

Ada beberapa operasi pengolahan citra yang biasanya dilakukan menurut jaim, 1989, diantaranya [7]:

2.2.2.1. Representasi dan pemodelan citra

Bertujuan untuk memberikan gambaran suatu hal. Contohnya adalah bagaimana pencahayaan benda disuatu lokasi dari hasil foto kamera, karakteristik jaringan tubuh dari citra X-ray, posisi target dari sebuah radar dari citra radar, suhu suatu daerah dari citra inframerah atau medan gravitasi suatu daerah dari citra geofisika [7].

2.2.2.2. Perbaikan citra

Bertujuan untuk menonjolkan fitur-fitur tertentu yang ada pada citra untuk kepentingan ekstraksi fitur, analisis citra atau menampilkan informasi citra tersebut. Contohnya adalah perbaikan kontras, perbaikan tepi objek, pemberian warna semu, penipisan derau, penajaman, dan sebagainya [7].

2.2.2.3. Pemugaran citra

Bertujuan untuk menghilangkan atau mengurangi degradasi citra. Mirip dengan perbaikan citra, hanya saja pada pemugaran citra sudah diketahui penyebab degradasi citra. Contohnya adalah menghilangkan kesamaran (*deblurring*) yang disebabkan oleh keterbatasan sensor, usia citra yang sudah tua atau banyaknya derau (*noise*) pada citra [7].

2.2.2.4. Analisis citra (*image analysis*)

Bertujuan untuk menghitung besaran kuantitatif dari citra untuk menghasilkan deskripsi dari citra tersebut. Teknik analisis citra memerlukan fitur-fitur tertentu untuk membantu dalam identifikasi objek. Proses segmentasi terkadang diperlukan untuk memisahkan objek yang diinginkan dari sekelilingnya sehingga selanjutnya dapat dilakukan pengukuran. Contohnya adalah mendeteksi tepi objek, ekstraksi batas (*boundary*), dan representasi daerah (*region*) [7].

2.2.2.5. Rekonstruksi citra

Bertujuan untuk mengurangi jumlah bit yang dibutuhkan untuk menyimpan atau mengirim gambar tanpa kehilangan informasi yang diperlukan. Citra berukuran besar akan mempengaruhi penggunaan memori dan lamanya waktu pengiriman [7].

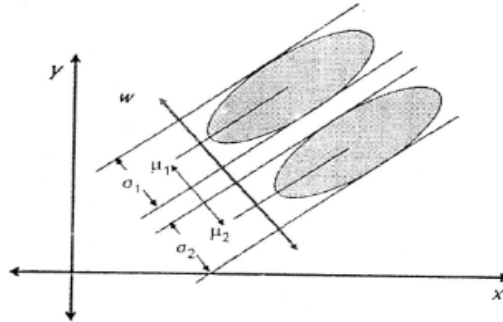
2.2.2.6. Kompresi citra

Bertujuan untuk mengurangi jumlah bit yang dibutuhkan untuk menyimpan atau mengirim gambar tanpa kehilangan informasi yang diperlukan. Citra yang berukuran besar akan mempengaruhi penggunaan memori dan lamanya waktu pengiriman.

2.2.3. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) merupakan metode klasik untuk analisis multivariat. Dimana bertujuan untuk mendapatkan cara yang efisien dalam menyajikan ulang ruang wajah dengan data pelatihan dibagi dalam beberapa kelas atau kategori wajah. Data gambar dalam data pelatihan dibagi ke dalam beberapa kelas sesuai dengan kelompoknya dengan menggunakan informasi ini *LDA* mempunyai ketepatan yang lebih tinggi dalam proses pengenalan data wajah dibandingkan dengan beberapa metode yang lain dalam ekstraksi ciri. Dimana

perbedaan dengan yang lain terdapat pada arah optimal vector w sehingga jarak antar *mean* maksimal dan lebar varian minimal tersebut dapat dilihat pada gambar berikut.



Gambar 2.1 Arah optimal vector w sehingga jarak antar mean maksimal dan lebar varian

a. Mencari matriks *scatter between class*

S_B disebut matriks *scatter between class* atau penyebaran data antar kelas yang berbeda. Bila pada PCA dicari rata-rata seluruh image saja, maka pada LDA juga harus dicari lebih dulu rata-rata image yang terdapat dalam satu kelas.

Berikut rumus mencari matriks *scatter between class* :

$$S_B = \sum_{i=1}^C N_i (\mu_i - \mu) (\mu_i - \mu)^T$$

Dimana N_i = Jumlah Image pada kelas ke i

μ_i = rata - rata kelas ke i

μ = rata-rata total

C = Jumlah kelas yang ada

b. Mencari *scatter within class*

S_W disebut matriks *scatter within class* atau penyebaran data dalam satu kelas yang sama. Berikut rumus mencari matriks *scatter within class* :

$$S_W = \sum_{i=1}^C N_i (X_k - \mu_i) (X_k - \mu_i)^T$$

Dimana X_k = Image ke k

C = Jumlah kelas yang ada

μ_i = rata-rata kelas ke i

c. Mencari Covarian matriks LDA

Berbeda dari PCA yang mendapatkan *covarian matrix* dari seluruh image dikurangi rata-rata totalnya, *covarian Matrix* LDA didapatkan dari operasi antara S_b dan S_w . dimana C adalah *covarian matrix* LDA. Maka Rumus dari *covarian matriks LDA* sebagai berikut :

$$C = S_B S_W^{-1}$$

d. Mencari Eigen vector dan Eigen value

Yang dimaksud dengan eigenvalue adalah sebuah bilangan skalar dan eigenvector adalah sebuah matriks yang keduanya dapat mendefinisikan matriks A .

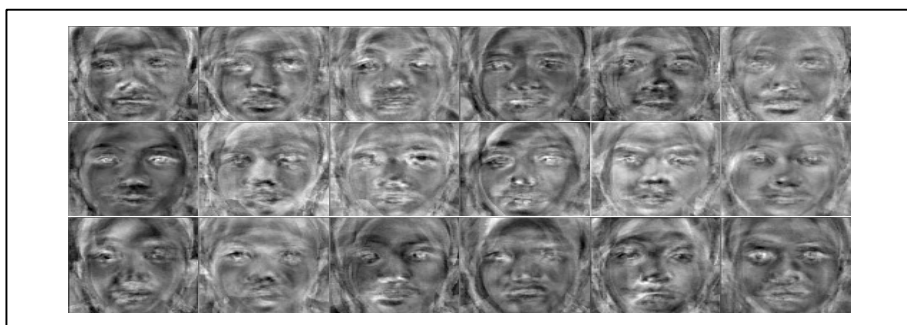
e. Mencari Feature LDA

Feature LDA dicari dengan cara yang sama seperti pada PCA. *Feature* dapat dicari dengan persamaan sebagai berikut

$$f = \sum_{i=1}^m (I_i - u)^T \times V$$

I = data tiap pixel dari image training ke- I

m = jumlah image training



Gambar 2.2 Hasil ekstraksi fitur citra data base

Gambar 2.2 merupakan hasil dari ekstraksi fitur citra *database* menggunakan metode LDA, walaupun hasil ekstraksi fitur tidak menyerupai gambar aslinya, akan tetapi setiap gambar yang merupakan satu kelas memiliki gambar yang hampir

sama yang diproyeksikan ke dalam kelas yang sama perbedaan gambarnya dapat dihilangkan atau diekstraksi dengan menjadi satu gambar.

2.2.4. Algoritma C4.5

Pada akhir tahun 1970 sampai awal tahun 1980 J. Ross Quinlan, seorang peneliti di bidang machine learning, membuat sebuah algoritma decision tree yang dikenal dengan ID3 (*Iterative Dichotomiser*). Quinlan kemudian membuat algoritma C4.5 (sering disebut dengan pohon keputusan) yang merupakan pengembangan dari algoritma ID3 [9]. Kelebihan dari algoritma ini yaitu mudah dimengerti, fleksibel dan menarik karena dapat divisualisasikan dalam bentuk gambar (pohon keputusan).

Secara umum algoritma C4.5 dapat dibangun menggunakan pohon keputusan dengan cara:

- a. Pilih atribut sebagai akar
- b. Buat cabang untuk masing-masing nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Tahapan dalam algoritma C4.5 ini adalah sebagai berikut:

1. Memilih atribut sebagai akar (*root*) dengan cara menghitung nilai entropy dan informasi gain dalam setiap atribut dengan nilai gain tertinggi menjadi akar pada sebuah pohon keputusan yang akan dibentuk. Adapun rumus tersebut adalah:

$$Entropy(S) = -\sum_{i=1}^n p_i * \log_2 p_i$$

Keterangan :

S = himpunan kasus

n = jumlah partisi S

p_i = proporsi dari S_i terhadap S

Sedangkan untuk mencari informasi gain menggunakan rumus sebagai berikut:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|s_i|}{|s|} * Entropy(S_i)$$

Keterangan :

S = Himpunan Kasus

A = Atribut

n = Jumlah partisi atribut A

$|s_i|$ = Jumlah kasus pada partisi ke-i

$|s|$ = Jumlah kasus dalam S

2. Langkah kedua adalah susun simpul berdasarkan atribut tersebut.
3. Bagi kasus-kasus yang dihasilkan kedalam cabang-cabang.
4. Ulangi proses perhitungan informasi gain diatas hingga semua data masuk kedalam kelas yang sama.

2.2.5. Penerapan Algoritma C4.5 Memprediksi Penjualan Wallpaper

Tabel 2.2 Data Wallpaper

NO	HARGA	UKURAN	JUMLAH MOTIF	KUALITAS BAHAN	WARNA	MINAT
1	Mahal	Besar	Sedikit	Baik	Banyak	Sedikit
2	Mahal	Kecil	Banyak	Terbaik	Banyak	Banyak
3	Murah	Kecil	Banyak	Baik	Banyak	Banyak
4	Mahal	Besar	Sedikit	Terbaik	Banyak	Sedikit
5	Murah	Besar	Sedikit	Terbaik	Sedikit	Sedikit
6	Murah	Besar	Sedikit	Terbaik	Banyak	Sedikit
7	Murah	Besar	Sedikit	Terbaik	Banyak	Sedikit
8	Murah	Kecil	Banyak	Terbaik	Banyak	Banyak
9	Mahal	Besar	Banyak	Terbaik	Banyak	Banyak

10	Mahal	Kecil	Banyak	Terbaik	Banyak	Banyak
11	Murah	Besar	Banyak	Baik	Banyak	Banyak
12	Mahal	Besar	Sedikit	Baik	Sedikit	Sedikit
13	Mahal	Kecil	Banyak	Terbaik	Sedikit	Sedikit
14	Mahal	Kecil	Sedikit	Terbaik	Sedikit	Sedikit
15	Mahal	Kecil	Banyak	Terbaik	Sedikit	Sedikit
16	Mahal	Kecil	Banyak	Baik	Sedikit	Sedikit
17	Mahal	Besar	Sedikit	Baik	Sedikit	Sedikit
18	Murah	Kecil	Sedikit	Baik	Sedikit	Sedikit
19	Murah	Kecil	Banyak	Baik	Sedikit	Banyak
20	Mahal	Besar	Banyak	Terbaik	Banyak	Banyak
21	Mahal	Besar	Sedikit	Baik	Banyak	Sedikit
22	Mahal	Kecil	Sedikit	Baik	Banyak	Sedikit
23	Mahal	Besar	Sedikit	Baik	Banyak	Sedikit
24	Mahal	Kecil	Banyak	Baik	Banyak	Banyak
25	Mahal	Besar	Banyak	Baik	Banyak	Banyak
26	Mahal	Kecil	Banyak	Baik	Banyak	Sedikit
27	Murah	Kecil	Banyak	Terbaik	Banyak	Sedikit
28	Murah	Besar	Banyak	Terbaik	Banyak	Banyak
29	Murah	Kecil	Banyak	Baik	Banyak	Sedikit
30	Murah	Besar	Banyak	Terbaik	Banyak	Banyak

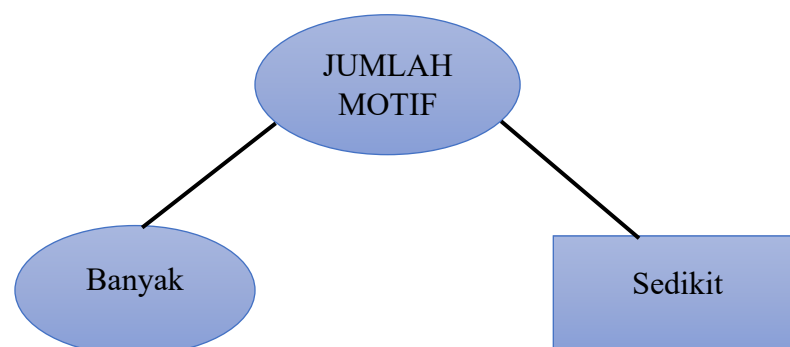
Pada table diatas terdapat 30 penilaian dari 6 penilaian yang menentukan minat konsumen terhadap wallpaper.

Tabel 2.3 Nilai Gain dan Nilai Entropy

NODE			JUMLAH	SEDIKIT	BANYAK	ENTROPY	GAIN
	TOTAL		30	18	12	0,970950554	
	HARGA						0,019975004
		MAHAL	18	12	0	0,418755834	
		MURAH	12	0	0	1	
	UKURAN						0
		BESAR	15	0	0	0,970556594	
		KECIL	15	0	0	0,836551554	
	JUMLAH MOTIF						0,419937094
		SEDIKIT	12	12	0	0	
		BANYAK	18	0	12	0,318295834	
	KUALITAS BAHAN						0,013406861
		TERBAIK	15	8	7	0,996795632	
		BAIK	15	10	5	0,838295834	
	WARNA						0,121118523
		BANYAK	21	10	11	0,998963633	
		SEDIKIT	9	8	8	0,583358335	

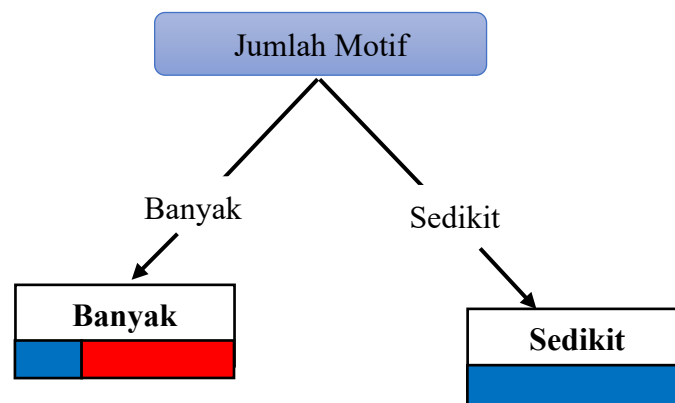
Dari tabel diatas dapat diketahui bahwa Gain tertinggi adalah jumlah motif yaitu sebesar 0,419973. Dengan demikian, jumlah motif dapat menjadi node akar.

Ada 2 nilai atribut dari jumlah motif, yaitu banyak dan sedikit, sehingga perlu dilakukan perhitungan lagi.

**Gambar 2.3** Decision Tree

Pada gambar diatas dapat diambil keputusan bahwa jika jumlah motif banyak maka minat konsumen banyak. Jika jumlah motif sedikit maka minat konsumen sedikit.

Setelah dilakukan dsengan perhitungan menggunakan excel, selanjutnya dilakukan perhitungan menggunakan aplikasi rapid miner. Setelah diproses maka didapatkan decision tree dibawah ini.



Tree

Jumlah Motif = Banyak; Banyak (Sedikit=6, Banyak=12)

Jumlah Motif = Sedikit; Sedikit (Sedikit=12, Banyak=0)

Gambar 2.4 Decision Tree dengan Rapid Miner

2.2.6 Evaluasi Model

Pada penelitian ini penulis menggunakan *Confusion Matrix* sebagai metode dalam perhitungan akurasi pada penerapan teknik pengenalan wajah menggunakan algoritma C4.5. *Confusion Matrix* adalah *tool* yang digunakan sebagai evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah matrix dari prediksi yang akan dibandingkan dengan kelas sebenarnya atau dengan kata lain berisi informasi nilai sebenarnya dan prediksi pada klasifikasi.

Rumusan ini menggunakan 4 keluaran yaitu :

1) *Recall*

Recall adalah proporsi kasus positif yang diidentifikasi dengan benar.

Rumus dari *Recall* = $d/(c+d)$

2) *Precision*

Precision adalah proporsi kasus dengan hasil positif yang benar.

Rumus dari *Precision* = $d/(b+d)$

3) *Accuracy*

Accuracy adalah perbandingan kasus yang diidentifikasi dengan benar dengan jumlah semua kasus.

Rumus dari *Accuracy* = $(a+c)/(a+b+c+d)$

4) *Error Rate*

Error Rate adalah kasus yang diidentifikasi salah dengan sejumlah semua kasus.

Rumus dari *Error Rate* = $(b+c)/(a+b+c+d)$

Keterangan:

a = jika hasil prediksi negatif dan data sebenarnya negatif.

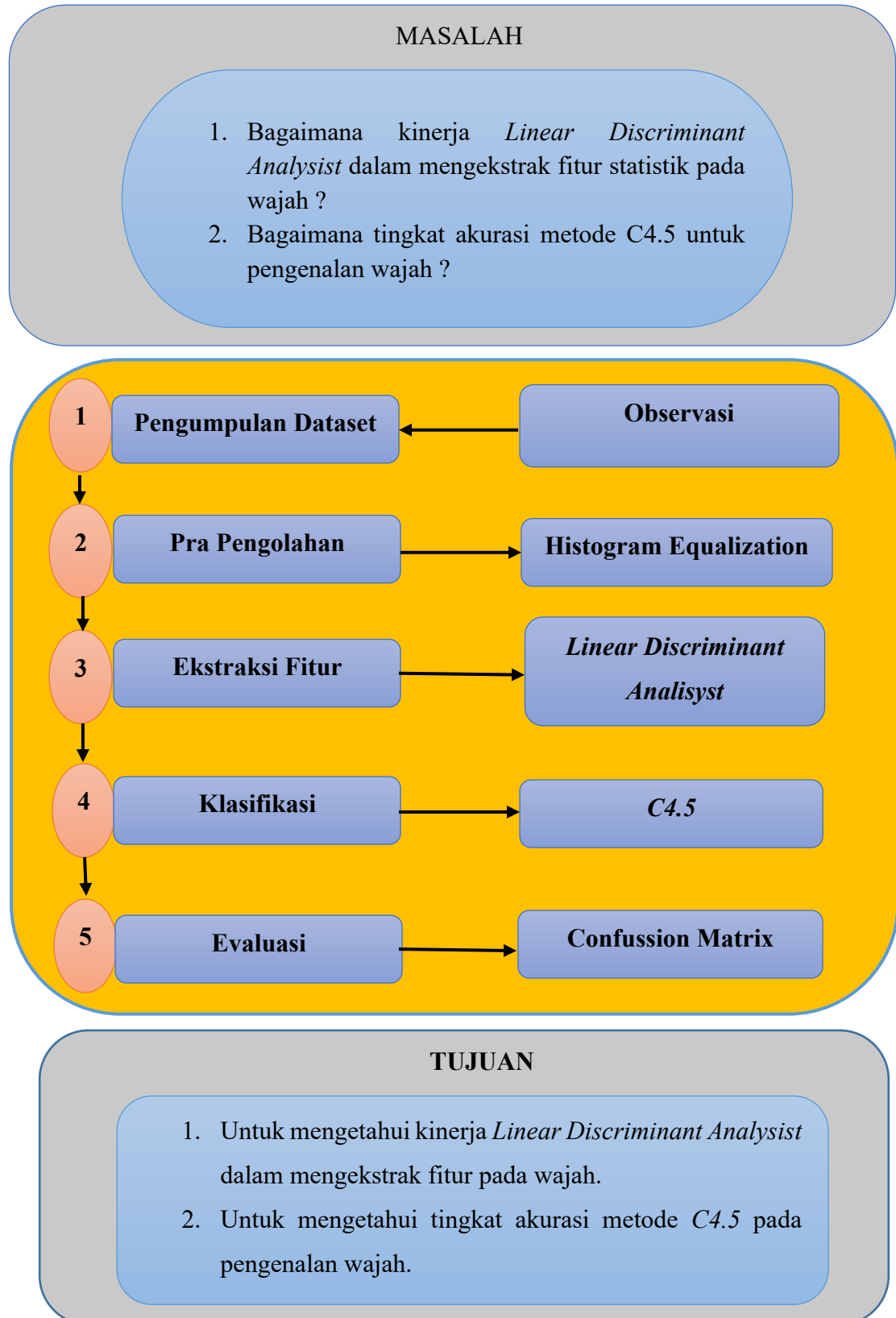
b = jika hasil prediksi positif sedangkan nilai sebenarnya negatif.

c = jika hasil prediksi negatif sedangkan nilai sebenarnya positif.

d = jika hasil prediksi positif sedangkan nilai sebenarnya positif.

2.3. Kerangka Pikir

Gambar 2.5 Bagan Kerangka Pikir



BAB III

METODE PENELITIAN

3.1 Jenis, Metode, Subjek, dan Objek

3.1.1. Jenis

Dilihat dari tingkat penerapannya, maka penelitian ini merupakan penelitian terapan.

3.1.2. Metode

Penelitian ini menggunakan metode penelitian eksperimen, dengan demikian jenis penelitian ini adalah penelitian eksperimental.

3.1.3. Subjek

Subjek penelitian ini adalah *Klasifikasi* menggunakan metode C4.5 untuk mendapatkan hasil akhir.

3.1.4. Objek

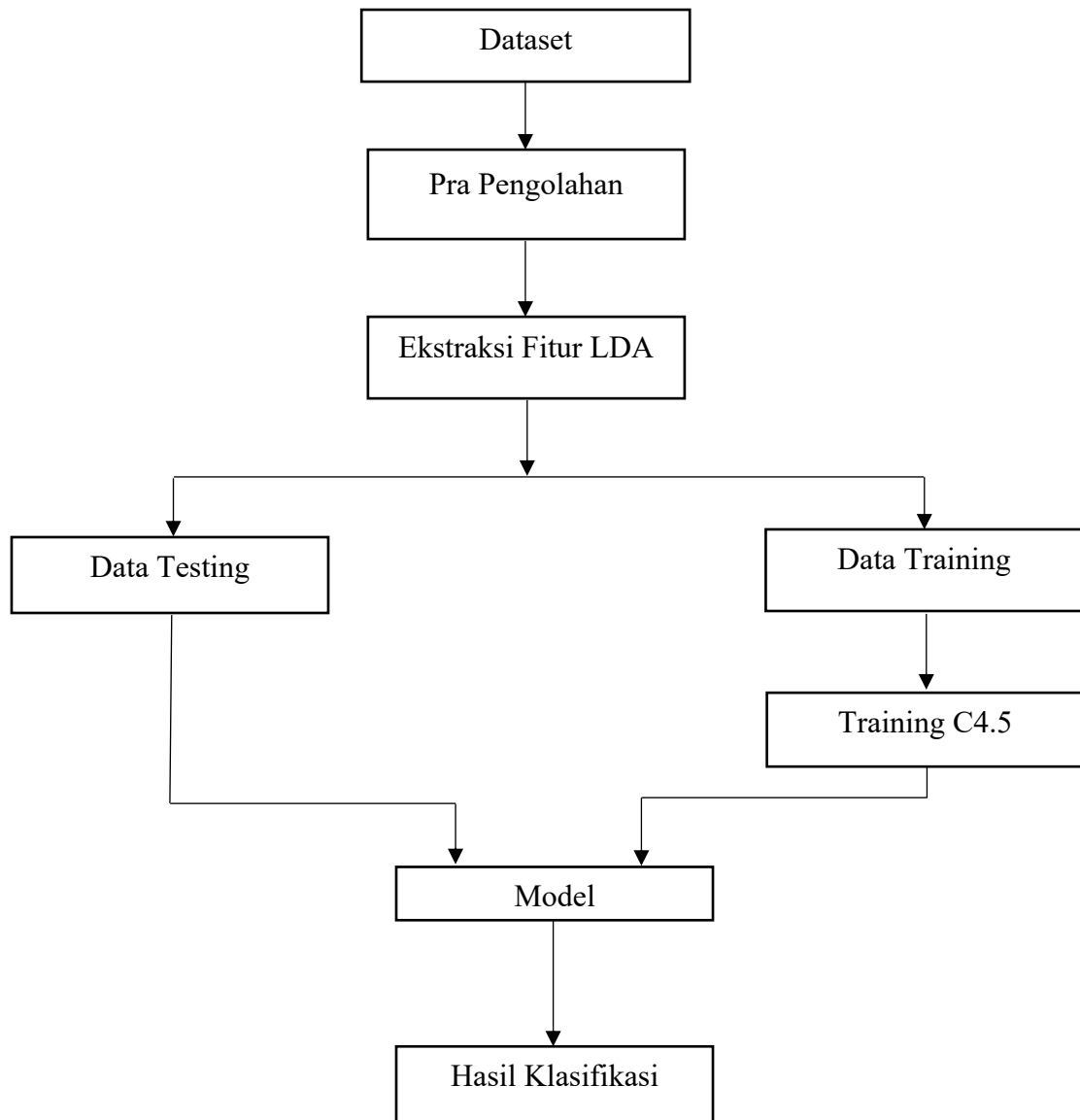
Objek yang digunakan pada penelitian ini adalah citra wajah manusia.

3.2 Pengumpulan Data

Data primer penelitian ini adalah pengambilan gambar dengan menggunakan teknik Observasi.

Dataset yang digunakan dalam penelitian ini adalah dataset publik yang di ambil dari penyedia dataset citra wajah yaitu *Original Images*, *dataset* ini di unduh dari <https://www.kaggle.com/datasets/vasukipatel/face-recognition-dataset/data?select=Original+Images>. Dataset ini merupakan data yang nantinya akan digunakan dalam proses fitur ekstraksi dan klasifikasi pengenalan wajah.

3.3 Pemodelan



Gambar 2.6 Pemodelan

3.3.1. Pra Pengolahan Data

Sebelum data diolah, terlebih dahulu dilakukan *Histogram Equalization*. *Histogram Equalization* adalah suatu proses untuk meratakan histogram agar derajat keabuan dari yang paling rendah (0) sampai dengan yang paling tinggi (255) mempunyai kemunculan yang rata, jadi dengan *Histogram Equalization* hasil gambar yang memiliki histogram yang tidak merata akan menjadi gambar yang lebih jelas karena derajat keabuannya tidak dominan gelap atau dominan terang.

3.3.2. Ekstraksi fitur Linear Discriminant Analysis (LDA)

Proses ekstraksi ini dilakukan untuk mengetahui fitur statistik pada wajah yang kemudian citra warna akan dirubah menjadi grayscale dan selanjutnya akan diklasifikasikan untuk hasil akhirnya.

3.3.3. Klasifikasi C4.5

Proses klasifikasi dilakukan untuk mengklasifikasikan setiap hasil akhir dari citra wajah. Kemudian akan dilihat berapa besar tingkat akurasi pengenalan wajah dengan klasifikasi C4.5.

3.3.4. Evaluasi Model

Model yang dilakukan kemudian dievaluasi dengan menggunakan *Confusion Matrix* untuk mengetahui tingkat akurasi.

BAB IV

HASIL PENELITIAN

4.1. Hasil Pengumpulan Data

Dataset yang digunakan pada penelitian ini terdiri dari data *public* dengan nama *Original Images*.



Gambar 4.1 Hasil Pengumpulan Dataset

Gambar diatas merupakan hasil pengumpulan dataset yang dilakukan dalam penelitian ini, dimana dataset yang digunakan berupa gambar wajah yang secara detail dijelaskan sebagai berikut;

Nama dataset : Original Images

Objek : wajah

Jenis data : Gambar

Format File : JPEG

Ukuran : bervariasi

Jumlah Objek : 31 orang

Total data : 315 gambar

Sumber data : <https://www.kaggle.com/datasets/vasukipatel/face-recognition-dataset/data?select=Original+Images>

4.2 Analisis dan Pemodelan

4.2.1. Pra Pengolahan Data

- Konversi gambar dari RGB ke Grayscale

Pada tahapan pra pengolahan yang dilakukan pada penelitian ini adalah dengan melakukan proses konversi gambar RGB yang diperoleh dari proses pengumpulan data gambar menjadi bentuk gray scale atau skala abu-abu citra, proses konversi gambar menjadi gray scale secara matematis dapat dilakukan dengan mencari nilai rata-rata untuk setiap channel gambar sehingga gambar yang dihasilkan nanti hanya tinggal 1 channel saja (1 layer).

$$g(x) = \frac{R + G + B}{3}$$

120	122	134	210	90	83
127	166	112	147	45	42
120	156	79	94	23	21
99	125	51	57	11	10
75	91	31	34	6	5
53	63	18	20	3	3

Channel R

105	107	119	195	75	68
112	151	97	132	30	27
105	141	64	79	38	36
84	110	36	42	26	25
60	76	16	19	21	20
38	48	3	5	18	18

Channel G

130	132	144	220	100	93
137	176	122	157	55	52
130	166	89	104	63	61
109	135	61	67	51	50
85	101	41	44	46	45
63	73	28	30	43	43

Channel B

Data *training* dan data *testing* yang awalnya berbentuk citra dari RGB (*red*, *green*, *blue*) dikonversi menjadi citra grayscale.

118	120	132	208	88	81
125	164	110	145	43	40
118	155	77	92	41	39
97	123	49	56	30	29
73	90	29	32	24	24
51	61	17	18	21	21

Gray Scale

- Peningkatan kualitas citra

Pada tahapan berikutnya pada proses pra pengolahan citra yang dilakukan pada penelitian ini adalah dengan melakukan proses peningkatan kualitas citra guna mendapatkan gambar atau citra yang memadai untuk proses lebih lanjut, proses ini perlu dilakukan karena pada saat dilakukan proses pengambilan gambar terdapat sejumlah gambar yang memiliki kualitas citra yang kurang memadai seperti gambar terlalu gelap atau terlalu terang, sehingga perlu dilakukan proses peningkatan citra guna menyeragamkan kondisi citra. Proses peningkatan kualitas citra sendiri pada penelitian ini dilakukan dengan cara operasi morfologi spasial pada setiap pixel citra dengan menggunakan kernel berukuran 3 x 3 dengan komposisi filter sebagai berikut:

0	-1	0
-1	5	-1
0	-1	0

Gambar 4.1.2 Filter 3 x 3

Proses peningkatan kualitas citra dilakukan dengan melakukan operasi morfologi citra terhadap gambar gray scale citra yang telah diperoleh dari tahap sebelumnya dengan memanfaatkan filter 3 x 3 diatas, secara matematis proses morfologi tersebut dapat dijelaskan sebagai berikut:

118	120	132	208	88	81
125	164	110	145	43	40
118	155	77	92	41	39
97	123	49	56	30	29
73	90	29	32	24	24
51	61	17	18	21	21

Gambar 4.1.3 Pixel Citra

0	-1	0
-1	5	-1
0	-1	0

Gambar 4.1.4 kernel/filter yang digunakan

Secara kerja proses morfologi dalam melakukan proses peningkatan kualitas citra adalah dengan melakukan proses perkalian antara kernel/filter yang ada dengan nilai pixel citra yang seukuran dengan ukuran filter yang digunakan yakni 3x3, proses perkalian tersebut dilakukan secara linear seperti yang dicontoh sebagai berikut:

$$g(x) = (0 * 118 + -1 * 120 + 0 * 132 + -1 * 125 + 5 * 164 + -1 * 110 + 0 * 118 + -1 * 155 + 0 * 77) / 9$$

$$g(x) = 35$$

sehingga hasil $g(x)$ tersebut dimasukkan kedalam matriks baru seperti ditampilkan pada gambar berikut:

118	120	132	208	88	81
125	35				40
118					39
97					29
73					24
51	61	17	18	21	21

Gambar 4.1.5 matriks hasil proses morfologi

Proses tersebut akan terus berulang hingga seluruh matriks tersebut terisi secara penuh seperti dibawah ini

118	120	132	208	88	81
125	35	37	30	11	40
118	32	40	29	18	39
97	42	26	34	12	29
73	45	45	23	13	24
51	61	17	18	21	21

Gambar 4.1.6 hasil akhir proses morfologi

4.2.2 Ekstraksi Linear Discriminant Analisis (LDA)

Linear Discriminant Analysis (LDA) adalah metode ekstraksi fitur dengan perpaduan dari perhitungan operasi matematika dan statistika yang diterapkan properti statistik terpisah untuk tiap obyek.

Mencari proyeksi terbaik yang dapat memisahkan data dengan pendekatan leastsquares. *Linear Discriminant Analysis* memberikan perlakuan statistik yang berbeda/terpisah untuk tiap-tiap kelas dengan menemukan kombinasi linier dari fitur yang menjadi ciri khas objek setiap kelas.

Linear Discriminant Analysis bertujuan untuk mencari proyeksi linear (*fisherface*) untuk memaksimalkan pemisahan antar kelas dan juga meminimumkan jarak didalam kelas objek yang sama.

118	120	132	208	88	81
125	35	37	30	11	40
118	32	40	29	18	39
97	42	26	34	12	29
73	45	45	23	13	24
51	61	17	18	21	21

Gambar 4.2.1 hasil normalisasi citra

Contoh dataset wajah yang sudah melalui proses peningkatan kualitas citra :

118	120	132	208	88	81	125	35	37	30	11	40	118	32	40	29	Akshay Kumar
80	156	123	100	34	35	78	88	28	18	98	76	115	54	51	20	Akhsay kumar
120	112	34	23	22	53	65	75	35	32	97	57	54	35	55	50	Akshay kumar
68	120	54	75	54	55	24	19	89	99	29	180	68	98	35	19	Alexandra Daddario
111	56	76	50	44	201	51	42	88	45	90	35	12	19	17	66	Alexandra Daddario
22	133	45	23	30	39	43	24	64	73	81	27	47	32	55	34	Alia Bhatt
85	91	209	17	94	19	29	19	95	29	58	100	82	71	23	31	Alia Bhatt
73	179	93	98	67	37	83	208	19	104	161	11	54	45	73	76	Alia Bhatt
53	34	64	42	23	204	144	32	53	64	32	36	23	64	20	19	Alia Bhatt
16	192	18	91	13	68	84	24	33	22	14	18	102	68	47	37	Amitabh Bachchan
93	38	82	80	30	94	81	104	11	99	39	21	60	87	27	19	Amitabh Bachchan
170	112	53	73	34	35	92	23	53	32	67	33	53	32	64	75	Amitabh Bachchan
102	69	98	82	18	20	24	45	82	92	78	82	19	38	109	91	Andy Samberg
192	180	90	82	37	89	20	74	92	47	93	84	39	29	103	64	Andy Samberg
28	92	102	22	118	21	12	79	39	38	77	37	83	13	20	88	Anushka Sharma
29	12	199	25	32	115	83	92	60	24	43	53	23	64	53	24	Anushka Sharma
182	23	53	75	65	65	39	24	69	39	58	46	86	55	122	35	Anushka Sharma

2. Memindahkan x berdasarkan group :

Memindahkan x berdasarkan group dengan memisahkan masing-masing nilai dari setiap data di setiap group yang berbeda.

X1	118	120	132	208	88	81	125	35	37	30	11	40	118	32	40	29
	80	156	123	100	34	35	78	88	28	18	98	76	115	54	51	20
	120	112	34	23	22	53	65	75	35	32	97	57	54	35	55	50
X2	68	120	54	75	54	55	24	19	89	99	29	180	68	98	35	19
	111	56	76	50	44	201	51	42	88	45	90	35	12	19	17	66
X3	22	133	45	23	30	39	43	24	64	73	81	27	47	32	55	34
	85	91	209	17	94	19	29	19	95	29	58	100	82	71	23	31
	73	179	93	98	67	37	83	208	19	104	161	11	54	45	73	76
	53	34	64	42	23	204	144	32	53	64	32	36	23	64	20	19
X4	16	192	18	91	13	68	84	24	33	22	14	18	102	68	47	37
	93	38	82	80	30	94	81	104	11	99	39	21	60	87	27	19
	170	112	53	73	34	35	92	23	53	32	67	33	53	32	64	75
X5	102	69	98	82	18	20	24	45	82	92	78	82	19	38	109	91
	192	180	90	82	37	89	20	74	92	47	93	84	39	29	103	64
X6	28	92	102	22	118	21	12	79	39	38	77	37	83	13	20	88
	29	12	199	25	32	115	83	92	60	24	43	53	23	64	53	24
	182	23	53	75	65	65	39	24	69	39	58	46	86	55	122	35

3. Hitung μ_i = mean feature dari group i dan μ = mean global

Setelah melakukan pemindahan nilai x di masing-masing group kemudian menghitung nilai μ_i dengan cara mengambil nilai rata-dari nilai x.

μ_1	106	129.33	96.333	110.33	48	56.333	89.333	66	33.333	26.667	68.667	57.667	95.667	40.333	48.667	33
μ_2	89.333	129.33	70.333	66	36.667	47.667	55.667	60.667	50.667	49.667	74.667	104.33	79	62.333	47	29.667
μ_3	99.667	96	54.667	49.333	40	103	46.667	45.333	70.667	58.667	72	90.667	44.667	50.667	35.667	45
μ_4	67	103	58.333	49.333	42.667	98.333	39.333	28.333	80.333	72.333	66.667	80.667	42.333	49.667	35.667	39.667
μ_5	72.667	93.333	110	30	56	86.333	41	28.333	82.333	49	76.333	54	47	40.667	31.667	43.667
μ_6	60	134.33	115.67	46	63.667	31.667	51.667	83.667	59.333	68.667	100	46	61	49.333	50.333	47

Nilai μ adalah rata-rata total

μ	90.706	101.12	89.706	68.588	47.235	72.412	63.353	59.235	55.706	52.176	66.235	55.059	61.059	49.176	53.765	45.706
-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

4. Hitung x_i^0 (Mean Corrected) :

(x_i minus mean global)

Mean Corrected data dihasilkan dari nilai x_i dikurang mean global

27.294	18.882	42.294	139.41	40.765	8.5882	61.647	-24.24	-18.71	-22.18	-55.24	-15.06	56.941	-17.18	-13.76	-16.71
-10.71	54.882	33.294	31.412	-13.24	-37.41	14.647	28.765	-27.71	-34.18	31.765	20.941	53.941	4.8235	-2.765	-25.71
29.294	10.882	-55.71	-45.59	-25.24	-19.41	1.6471	15.765	-20.71	-20.18	30.765	1.9412	-7.059	-14.18	1.2353	4.2941
-22.71	18.882	-35.71	6.4118	6.7647	-17.41	-39.35	-40.24	33.294	46.824	-37.24	124.94	6.9412	48.824	-18.76	-26.71
20.294	-45.12	-13.71	-18.59	-3.235	128.59	-12.35	-17.24	32.294	-7.176	23.765	-20.06	-49.06	-30.18	-36.76	20.294
-68.71	31.882	-44.71	-45.59	-17.24	-33.41	-20.35	-35.24	8.2941	20.824	14.765	-28.06	-14.06	-17.18	1.2353	-11.71
-5.706	-10.12	119.29	-51.59	46.765	-53.41	-34.35	-40.24	39.294	-23.18	-8.235	44.941	20.941	21.824	-30.76	-14.71
-17.71	77.882	3.2941	29.412	19.765	-35.41	19.647	148.76	-36.71	51.824	94.765	-44.06	-7.059	-4.176	19.235	30.294
-37.71	-67.12	-25.71	-26.59	-24.24	131.59	80.647	-27.24	-2.706	11.824	-34.24	-19.06	-38.06	14.824	-33.76	-26.71
-74.71	90.882	-71.71	22.412	-34.24	-4.412	20.647	-35.24	-22.71	-30.18	-52.24	-37.06	40.941	18.824	-6.765	-8.706
2.2941	-63.12	-7.706	11.412	-17.24	21.588	17.647	44.765	-44.71	46.824	-27.24	-34.06	-1.059	37.824	-26.76	-26.71
79.294	10.882	-36.71	4.4118	-13.24	-37.41	28.647	-36.24	-2.706	-20.18	0.7647	-22.06	-8.059	-17.18	10.235	29.294
11.294	-32.12	8.2941	13.412	-29.24	-52.41	-39.35	-14.24	26.294	39.824	11.765	26.941	-42.06	-11.18	55.235	45.294
101.29	78.882	0.2941	13.412	-10.24	16.588	-43.35	14.765	36.294	-5.176	26.765	28.941	-22.06	-20.18	49.235	18.294
-62.71	-9.118	12.294	-46.59	70.765	-51.41	-51.35	19.765	-16.71	-14.18	10.765	-18.06	21.941	-36.18	-33.76	42.294
-61.71	-89.12	109.29	-43.59	-15.24	42.588	19.647	32.765	4.2941	-28.18	-23.24	-2.059	-38.06	14.824	-0.765	-21.71
91.294	-78.12	-36.71	6.4118	17.765	-7.412	-24.35	-35.24	13.294	-13.18	-8.235	-9.059	24.941	5.8235	68.235	-10.71

5. Hitung matriks kovarian grup i

$$C_i = \frac{(x_i^0)^T x_i^0}{n_i} \text{ matriks kovarian grup i}$$

572,5767	82,200692	-277,9723	711,12111	171,6955	22,094579	524,68051	-169,2065	-273,5017	-276,8166	-315,481	-192,782	256,62976	-311,9146	-103,3033	-18,32526
82,200692	1162,3472	673,2203	1286,7555	-77,08997	-700,7751	661,94348	430,87082	-699,7013	-837,9988	345,04729	295,35986	1319,5952	-71,29296	-132,7336	-559,8385
-277,9723	673,2203	2000,1453	3160,5525	896,40138	66,329873	1001,0727	-315,1672	-186,7174	-317,2872	-997,4418	-15,93887	1532,4729	74,614764	-247,6759	-600,5409
711,12111	1286,7555	3160,5525	7500,2088	2139,2561	302,35986	2993,1096	-1064,607	-844,722	-1081,798	-2701,724	-510,0242	3318,1522	-532,2687	-687,3737	-1110,742
171,6955	-77,08997	896,40138	2139,2561	824,58478	445,03806	759,20069	-588,827	42,224913	19,15917	-1149,474	-313,3391	595,13149	-135,4291	-185,2318	-149,7163
22,094579	-700,7751	66,329873	302,35986	445,03806	616,73818	-16,83506	-530,0992	425,93772	493,26874	-753,3149	-316,8189	-463,9954	-17,594	-12,92042	244,95732
524,68051	661,94348	1001,0727	2993,1096	759,20069	-16,83506	1339,203	-348,917	-531,0254	-633,6436	-963,0542	-206,1361	1429,5698	-337,1926	-295,6713	-466,4371
-169,2065	430,87082	-315,1672	-1064,607	-588,827	-530,0992	-348,917	554,42791	-223,3437	-254,5663	912,44752	332,64129	20,11188	110,51211	91,179931	-88,95156
-273,5017	-699,7013	-186,7174	-844,722	42,224913	425,93772	-531,0254	-223,3437	515,41984	593,16378	-161,2849	-112,8997	-804,4879	160,39908	102,83391	311,92964
-276,8166	-837,9988	-317,2872	-1081,798	19,15917	493,26874	-633,6436	-254,5663	593,16378	688,97232	-160,4683	-140,3033	-987,9504	167,36448	124,93887	387,4579
-315,481	345,04729	-997,4418	-2701,724	-1149,474	-753,3149	-963,0542	912,44752	-161,2849	-160,4683	1668,8005	518,89619	-549,6332	221,94348	236,82699	79,4406
-192,782	295,35986	-15,93887	-510,0242	-313,3391	-316,8189	-206,1361	332,64129	-112,8997	-140,3033	518,89619	223,02307	86,140715	110,71626	50,594002	-92,80161
256,62976	1319,5952	1532,4729	3318,1522	595,13149	-463,9954	1429,5698	20,11188	-804,4879	-987,9504	-549,6332	86,140715	2067,2584	-205,9308	-313,8766	-789,3899
-311,9146	-71,29296	74,614764	-532,2687	-135,4291	-17,594	-337,1926	110,51211	160,39908	167,36448	221,94348	110,71626	-205,9308	173,08997	68,527105	34,026528
-103,3033	-132,7336	-247,6759	-687,3737	-185,2318	-12,92042	-295,6713	91,179931	102,83391	124,93887	236,82699	50,594002	-313,8766	68,527105	66,212226	102,10842
-18,32526	-559,8385	-600,5409	-1110,742	-149,7163	244,95732	-466,4371	-88,95156	311,92964	387,4579	79,4406	-92,80161	-789,3899	34,026528	102,10842	319,43945

$$C(r, s) = \frac{1}{n} \sum_{i=1}^g n_i \cdot c_i(r, s)$$

=Dikumpulkan dalam matriks kovarians group. Ini dihitung dalam setiap entri (r,s) dalam matriks. Dalam contoh ini, $3/17*572.5767+2/17*267586.3+4/17*1622.072+3/17*3957.93+2/17*3957.93+3/17*5194.028 = 34218.512$

6. Menghitung Pooled Wirthin Group

Menghitung Pooled Wirthin Group dengan cara mengalikan nilai dari Covarian masing-masing grup dengan nilai Prior Probabiliti dari masing-masing grup.

C =

34218.5	53033.2	140918	336077	97827.7	16951.6	132607	-50759	-34527	-45140	-124362	-24076	144812	-23829	-29684	-47723
53033.2	100624	234439	558754	158137	19918.3	222954	-77391	-64332	-82106	-198544	-37278	249037	-40018	-51067	-83052
140918	234439	637388	1507133	441853	79515.7	596332	-228649	-154480	-200387	-562991	-110710	648327	-105918	-137899	-214608
336077	558754	1507133	3580205	1047695	189438	1416720	-543638	-367690	-474852	-1E+06	-264627	1539476	-251933	-326338	-508949
97827.7	158137	441853	1047695	310027	59208.2	413197	-162411	-104216	-135325	-395665	-79327	446803	-73711	-95691	-146901
16951.6	19918.3	79515.7	189438	59208.2	19017	74086.4	-34527	-13963	-18675	-78863	-17458	74137.4	-12687	-17391	-24228
132607	222954	596332	1416720	413197	74086.4	562129	-213665	-147274	-189657	-527471	-104214	610870	-99551	-129693	-202657
-50759	-77391	-228649	-543638	-162411	-34527	-213665	89128.6	50876.7	67631.7	210074	42478.8	-228581	38030.2	49506.7	74988.7
-34527	-64332	-154480	-367690	-104216	-13963	-147274	50876.7	42638.6	53716.6	131347	24989.2	-163653	26050.4	33961.4	54905.4
-45140	-82106	-200387	-474852	-135325	-18675	-189657	67631.7	53716.6	69577	170967	31912.3	-210881	33766.9	43650.1	70678
-124362	-198544	-562991	-1E+06	-395665	-78863	-527471	210074	131347	170967	508300	102308	-567794	93503.8	122125	187074
-24076	-37278	-110710	-264627	-79327	-17458	-104214	42478.8	24989.2	31912.3	102308	21710.7	-110577	18451.7	24175.9	36099.7
144812	249037	648327	1539476	446803	74137.4	610870	-228581	-163653	-210881	-567794	-110577	669365	-108625	-140838	-222224
-23829	-40018	-105918	-251933	-73711	-12687	-99551	38030.2	26050.4	33766.9	93503.8	18451.7	-108625	18079.6	22856.9	35675.4
-29684	-51067	-137899	-326338	-95691	-17391	-129693	49506.7	33961.4	43650.1	122125	24175.9	-140838	22856.9	30681.7	46811.2
-47723	-83052	-214608	-508949	-146901	-24228	-202657	74988.7	54905.4	70678	187074	36099.7	-222224	35675.4	46811.2	74441.9

Invers matriks yang di kumpulkan adalah

$$C^{-i} =$$

-2E+07	75211778	36917311	75437025	2.31E+08	39771078	-1E+08	-2.3E+08	-5E+08	-1E+08	1.28E+08	53469639	-3.7E+08	2.28E+08	1.19E+08	-5.8E+07
75211778	-2.8E+08	-1.4E+08	-2.8E+08	-8.5E+08	-1.5E+08	3.85E+08	8.57E+08	1.84E+09	3.73E+08	-4.7E+08	-2E+08	1.36E+09	-8.4E+08	-4.4E+08	2.12E+08
36917311	-1.4E+08	-6.7E+07	-1.4E+08	-4.2E+08	-7.2E+07	1.89E+08	4.21E+08	9.01E+08	1.83E+08	-2.3E+08	-9.6E+07	6.65E+08	-4.1E+08	-2.2E+08	1.04E+08
75437025	-2.8E+08	-1.4E+08	-2.8E+08	-8.5E+08	-1.5E+08	3.86E+08	8.6E+08	1.84E+09	3.74E+08	-4.7E+08	-2E+08	1.36E+09	-8.4E+08	-4.4E+08	2.13E+08
2.31E+08	-8.5E+08	-4.2E+08	-8.5E+08	-2.6E+09	-4.5E+08	1.18E+09	2.63E+09	5.64E+09	1.15E+09	-1.4E+09	-6E+08	4.16E+09	-2.6E+09	-1.3E+09	6.52E+08
39771078	-1.5E+08	-7.2E+07	-1.5E+08	-4.5E+08	-7.7E+07	2.04E+08	4.53E+08	9.71E+08	1.97E+08	-2.5E+08	-1E+08	7.17E+08	-4.4E+08	-2.3E+08	1.12E+08
-1E+08	3.85E+08	1.89E+08	3.86E+08	1.18E+09	2.04E+08	-5.4E+08	-1.2E+09	-2.6E+09	-5.2E+08	6.55E+08	2.74E+08	-1.9E+09	1.16E+09	6.11E+08	-3E+08
-2.3E+08	8.57E+08	4.21E+08	8.6E+08	2.63E+09	4.53E+08	-1.2E+09	-2.7E+09	-5.7E+09	-1.2E+09	1.46E+09	6.09E+08	-4.2E+09	2.59E+09	1.36E+09	-6.6E+08
-5E+08	1.84E+09	9.01E+08	1.84E+09	5.64E+09	9.71E+08	-2.6E+09	-5.7E+09	-1.2E+10	-2.5E+09	3.13E+09	1.31E+09	-9E+09	5.56E+09	2.91E+09	-1.4E+09
-1E+08	3.73E+08	1.83E+08	3.74E+08	1.15E+09	1.97E+08	-5.2E+08	-1.2E+09	-2.5E+09	-5E+08	6.35E+08	2.65E+08	-1.8E+09	1.13E+09	5.92E+08	-2.9E+08
1.28E+08	-4.7E+08	-2.3E+08	-4.7E+08	-1.4E+09	-2.5E+08	6.55E+08	1.46E+09	3.13E+09	6.35E+08	-8E+08	-3.3E+08	2.31E+09	-1.4E+09	-7.5E+08	3.61E+08
53469639	-2E+08	-9.6E+07	-2E+08	-6E+08	-1E+08	2.74E+08	6.09E+08	1.31E+09	2.65E+08	-3.3E+08	-1.4E+08	9.64E+08	-5.9E+08	-3.1E+08	1.51E+08
-3.7E+08	1.36E+09	6.65E+08	1.36E+09	4.16E+09	7.17E+08	-1.9E+09	-4.2E+09	-9E+09	-1.8E+09	2.31E+09	9.64E+08	-6.7E+09	4.1E+09	2.15E+09	-1E+09
2.28E+08	-8.4E+08	-4.1E+08	-8.4E+08	-2.6E+09	-4.4E+08	1.16E+09	2.59E+09	5.56E+09	1.13E+09	-1.4E+09	-5.9E+08	4.1E+09	-2.5E+09	-1.3E+09	6.42E+08
1.19E+08	-4.4E+08	-2.2E+08	-4.4E+08	-1.3E+09	-2.3E+08	6.11E+08	1.36E+09	2.91E+09	5.92E+08	-7.5E+08	-3.1E+08	2.15E+09	-1.3E+09	-7E+08	3.37E+08
-5.8E+07	2.12E+08	1.04E+08	2.13E+08	6.52E+08	1.12E+08	-3E+08	-6.6E+08	-1.4E+09	-2.9E+08	3.61E+08	1.51E+08	-1E+09	6.42E+08	3.37E+08	-1.6E+08

$\mathbf{P}$ = vektor probabilitas sebelumnya (setiap baris mewakili probabilitas sebelumnya dari grup i) jika kita tidak mengetahui probabilitas priornya, kita asumsikan saja sama dengan total sampel tiap kelompok dibagi total sampel, ya itu

$$p_i = \frac{n_i}{N} = p_i = \frac{3}{17}$$

Prior Probabiliti

Grup 1	0.176471
Grup 2	0.117647
Grup 3	0.235294
Grup 4	0.176471
Grup 5	0.117647
Grup 6	0.176471

Fungsi diskriminant

$$f_i = \mu_i \mathbf{C}^{-1} \mathbf{x}_k^T - \frac{1}{2} \mu_i \mathbf{C}^{-1} \mu_i^T + \ln(p_i)$$

Dan didapatkan nilai fungsi diskriminantnya adalah

Nilai F1 = -1.15282E+13

Nilai F2 = -1.15282E+13

Nilai F3 = -1.15282E+13

Nilai F4 = -7.70559E+12

4.2.3 Klasifikasi C4.5

Algoritma C4.5 atau pohon keputusan (*Decision Tree*) merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variable input menjadi variable target. Karena pohon keputusan memadukan antara eksplorasi

data dan pemodelan, sangat bagus sekali sebagai langkah awal dalam proses pemodelan bahkan ketika diadakan sebagai model akhir dari beberapa teknik lain.

Pohon keputusan membangun model klasifikasi atau regresi dalam bentuk struktur pohon. Ini memecah kumpulan data menjadi subkumpulan yang lebih kecil dan pada saat yang sama pohon keputusan terkait dikembangkan secara bertahap. Pohon keputusan dapat menangani data kategorikal dan numerik.

Tabel 4.1.Hasil Ekstraksi Fitur

NAMA	F1	F2	F3	F4
Akshay Kumar	0.420497761	-0.315021408	-0.349097413	-0.393039029
	0.006016133	-0.466228326	-1.311326425	-0.763179043
	0.920232157	-1.020685183	0.432155544	-0.484971377
Alexandra	-0.809668797	-1.322446458	-2.88809915	2.155076139
	-2.576526644	0.378414258	-1.683649597	-0.333073587
Alia Bhatt	-1.555172917	1.459099009	5.050273806	1.947934469
	-1.725299937	1.146888813	3.077494937	2.113308706
	-1.631738846	2.563100025	3.827775558	-1.050998414
	0.30084463	1.864474156	4.145697656	0.362089021
Amitabh	-0.026188511	0.395813851	0.507928326	-0.433356998
	-0.747518047	-1.536755192	2.061957528	-0.718927076
	-0.234114663	-0.858598152	0.455876611	0.160388379
Andy Samb	-0.044311639	0.715367315	0.853587378	-0.094275549
	-0.14717571	0.2929482	-0.66327768	0.565713184
Anushka	-1.378303875	-0.418712754	-0.591908612	-0.88980856
	-2.195894755	1.687125221	2.165514478	0.851168033
	-1.049894807	-0.032348061	1.386748376	-0.493099995

Kemudian hitung entropy dari data tabel keseluruhan dataset diatas sebagai berikut:

$$\begin{aligned} \text{Jadi } Entropy(S) = & (-10/251 * \text{LOG}(2,10/251) + -8/251 * \text{LOG}(2,8/251) + - \\ & 12/251 * \text{LOG}(2,12/251) + -7/251 * \text{LOG}(2,7/251) + -9/251 * \text{LOG}(2,9/251) + - \\ & 10/251 * \text{LOG}(2,10/251) + -8/251 * \text{LOG}(2,8/251) + -10/251 * \text{LOG}(2,10/251) + - \\ & 7/251 * \text{LOG}(2,7/251) + -11/251 * \text{LOG}(2,11/251) + -10/251 * \text{LOG}(2,10/251) + - \\ & 9/251 * \text{LOG}(2,9/251) + -7/251 * \text{LOG}(2,7/251) + -14/251 * \text{LOG}(2,14/251) + - \\ & 11/251 * \text{LOG}(2,11/251) + -8/251 * \text{LOG}(2,8/251) + -10/251 * \text{LOG}(2,10/251) + - \\ & 10/251 * \text{LOG}(2,10/251) + -5/251 * \text{LOG}(2,5/251) + -6/251 * \text{LOG}(2,6/251) + - \\ & 10/251 * \text{LOG}(2,10/251) + -5/251 * \text{LOG}(2,5/251) + -10/251 * \text{LOG}(2,10/251) + - \\ & 6/251 * \text{LOG}(2,6/251) + -5/251 * \text{LOG}(2,5/251) + -9/251 * \text{LOG}(2,9/251) + - \\ & 3/251 * \text{LOG}(2,3/251) + -7/251 * \text{LOG}(2,7/251) + -4/251 * \text{LOG}(2,4/251) + - \\ & 5/251 * \text{LOG}(2,5/251) + -5/251 * \text{LOG}(2,5/251) = 0.20655002 \end{aligned}$$

Tabel 4.2 Hasil Perhitungan pada *Dataset*

Total	Akshay Kumar	Alexandra	Alia Bhatt	Amitabh	Entropy
251	0.848283149	0.753951562	0.841868304	0.5511716	0.20655002

Tabel 4.3 Analisis Atribut, Nilai, Banyaknya Kejadian Nilai, Entropy dan Gain

	Akshay Kumar	Alexandra	Alia Bhatt	Amitabh	Entropy	Entropy	Entropy	Entropy	Entropy(S)	GAIN
f1	0.848283149	0.753951562	0.841868304	0.5511716	3.57349111	1.850386524	3.390071048	0.641324853	0.20655002	-15.75366322
f2	1.289565685	0.675993785	0.655444655	0.744216862	-3.514901869	1.196622609	1.075461833	1.746147589	0.20655002	-8.319559118
f3	1.326975983	0.797582973	1.65807305	1.034499927	-3.251258483	2.444372979	-2.272866136	-21.14091568	0.20655002	-893.2453579
f4	1.066767354	1.038176013	1.376070063	0.602656994	11.4404066	-19.20732019	2.987858711	-0.824889728	0.20655002	-1806.547546

Dari hasil pada table diatas, dapat diketahui bahwa attribut dengan Gain tertinggi adalah F4 yaitu sebesar -1806.547546. Dengan demikian F4 dapat menjadi node akar.

Berikut ini adalah hasil pengujian data *testing* menggunakan fitur ekstraksi *Linear Discriminant Analysis* dan metode klasifikasi *Algoritma C4.5*. Peneliti memasukan target dan prediksi. Target adalah data yang diharapkan menjadi hasil yang telah di prediksi sedangkan prediksi memperkirakan hasil yang akan di hasilkan. Prediksi yang dihasilkan banyak yang tidak sesuai dengan target karena ada beberapa hasil yang tidak sesuai.

Tabel.4.6 Hasil perhitungan C4.5 untuk target dan prediksi

No	Target	Prediksi
1	Priyanka Chopra	Camila Cabello
2	Marmik	Jessica Alba
3	Tom Cruise	Tom Cruise
4	Andy Samberg	Jessica Alba
5	Zac Efron	Akshay Kumar
6	Amitabh Bachchan	Alexandra Daddario
7	Alexandra Daddario	Hrithik Roshan
8	Courtney Cox	Courtney Cox
9	Amitabh Bachchan	Akshay Kumar
10	Zac Efron	Jessica Alba
11	Andy Samberg	Camila Cabello
12	Alexandra Daddario	Alexandra Daddario
13	Lisa Kudrow	Tom Cruise
14	Tom Cruise	Henry Cavill
15	Andy Samberg	Jessica Alba
16	Natalie Portman	Jessica Alba
17	Alexandra Daddario	Hrithik Roshan
18	Hrithik Roshan	Akshay Kumar
19	Jessica Alba	Kashyap
20	Hrithik Roshan	Roger Federer
21	Lisa Kudrow	Margot Robbie
22	Alexandra Daddario	Roger Federer
23	Akshay Kumar	Hrithik Roshan
24	Zac Efron	Lisa Kudrow
25	Robert Downey Jr	Courtney Cox
26	Elizabeth Olsen	Roger Federer
27	Hugh Jackman	Jessica Alba
28	Marmik	Amitabh Bachchan
29	Henry Cavill	Hrithik Roshan
30	Robert Downey Jr	Jessica Alba

31	Billie Eilish	Jessica Alba
32	Elizabeth Olsen	Courtney Cox
33	Billie Eilish	Alexandra Daddario
34	Hugh Jackman	Courtney Cox
35	Alexandra Daddario	Hrithik Roshan
36	Brad Pitt	Alia Bhatt
37	Margot Robbie	Margot Robbie
38	Billie Eilish	Dwayne Johnson
39	Claire Holt	Lisa Kudrow
40	Alexandra Daddario	Alexandra Daddario
41	Courtney Cox	Courtney Cox
42	Charlize Theron	Claire Holt
43	Dwayne Johnson	Tom Cruise
44	Ellen Degeneres	Margot Robbie
45	Lisa Kudrow	Margot Robbie
46	Elizabeth Olsen	Alexandra Daddario
47	Ellen Degeneres	Margot Robbie
48	Zac Efron	Amitabh Bachchan
49	Jessica Alba	Courtney Cox
50	Courtney Cox	Jessica Alba
51	Andy Samberg	Jessica Alba
52	Natalie Portman	Jessica Alba
53	Natalie Portman	Hrithik Roshan
54	Zac Efron	Jessica Alba
55	Charlize Theron	Margot Robbie
56	Henry Cavill	Margot Robbie
57	Vijay Deverakonda	Roger Federer
58	Virat Kohli	Hrithik Roshan
59	Dwayne Johnson	Amitabh Bachchan
60	Camila Cabello	Alexandra Daddario
61	Priyanka Chopra	Jessica Alba
62	Ellen Degeneres	Lisa Kudrow
63	Natalie Portman	Jessica Alba

4.3. Confusion Matrix

Pada penelitian ini menggunakan Confusion Matrix sebagai metode dalam perhitungan akurasi pada penerapan pengenalan wajah seseorang. Confision Matrix dapat diartikan sebagai suatu alat yang memiliki fungsi untuk melakukan analisis

apakah classifier tersebut baik dalam mengenali tuple dari kelas yang berbeda. Nilai dari TruePositive dan True-Negative memberikan informasi ketika classifier dalam melakukan klasifikasi data bernilai benar, sedangkan FalsePositive dan False-Negative memberikan informasi ketika classifier salah dalam melakukan klasifikasi data.

Jumlah data yang diklasifikasi dengan benar = 6

Jumlah data yang diklasifikasi salah = 57

$$\text{Akurasi} = \frac{\text{Data Uji Dikenali}}{\text{Data Uji}} \times 100 \% \quad \text{Akurasi} = \frac{6}{63} \times 100 \% = 9.5 \%$$

BAB V

PEMBAHASAN

5.1 Objek penelitian

Pengumpulan data yang diambil dari penyedia dataset <http://www.kaggle.com/datasets/vasukipatel/face-recognition-dataset/data?select=Original+Images> dengan jumlah data sebanyak 2.562 *image*, beberapa *image* tidak bisa diolah karena faktor blur, objek wajah yang miring dan tidak bisa dideteksi wajahnya, sehingga peneliti mengambil data sebanyak 315 dari 2.562 *image* yang dipilih berdasarkan hasil prapengolahan sistem. 315 *image* yang dipakai dalam penelitian ini terdiri dari 252 data *training* dan 63 data *testing* yang merupakan data *image RGB* dengan format *JPEG*.

5.2 Pembahasan Model

Berdasarkan perolehan hasil penelitian yang telah dilakukan pada bab sebelumnya, maka dapat dijelaskan beberapa hal sebagai berikut :

- Data gambar yang berhasil dihasilkan pada penelitian ini adalah 315 gambar wajah, yang dibagi dalam 31 folder. Berdasarkan visualisasi citra bahwa keseluruhan citra dapat dibedakan dari wajah masing-masing nama sesuai nama folder.
- Pada percobaan yang dilakukan pada penelitian ini fitur citra di ekstraksi menggunakan metode *Linear Discriminant Analysis*. Sebelum pengolahan data dilakukan terlebih dahulu peneliti melakukan pengolahan data yaitu mengkonversi data *training* dan data *testing* dari *image RGB* ke *Grayscale*. Hal ini dilakukan karena inputan dari metode *Linear Discriminant Analysis* merupakan inputan *image grayscale*.
- Setelah pohon keputusan terbentuk sepenuhnya, algoritma C4.5 dapat melakukan pemangkasan terhadap nilai Gain terbesar yang dijadikan Node akar.

- Berdasarkan hasil proses klasifikasi yang dilakukan terhadap data testing yang digunakan berjumlah sebanyak 63 gambar, dapat diperoleh nilai kinerja model dari algoritma C4.5 yang digunakan yakni nilai data yang benar sebanyak 6 data gambar, dan nilai data yang salah sebanyak 57 data gambar, dan nilai Akurasi data testing hanya berkisar diangka 9.5 %. Sehingga berdasarkan perolehan nilai kinerja model tersebut menunjukkan bahwa penggunaan algoritma C4.5 tidak dapat bekerja dengan baik, sehingga penggunaan metode tersebut membutuhkan metode ekstraksi selain Linear Discriminant Analysis untuk menghasilkan nilai akurasi yang cukup baik kedepannya.

BAB VI

PENUTUP

6.1.Kesimpulan

Berdasarkan hasil penelitian yang dilakukan pada pengenalan wajah dan pembahasan yang telah diuraikan sebelumnya maka dapat ditarik suatu kesimpulan bahwa:

1. Hasil uji coba metode *Linear Discriminant Analysis* pada pengenalan wajah mendapatkan hasil yang sangat rendah dan kurang baik.
2. Dapat menggunakan metode *C4.5* untuk Klasifikasi, mendapatkan tingkat akurasi sebesar 9.5% karena metode *Linear Discriminant Analysis* tidak cocok dikombinasikan dengan Algoritma *C4.5*.

6.2. Saran

Setelah melakukan Penelitian Reduksi Fitur pada pengenalan wajah, ada beberapa saran yang perlu di perhatikan untuk mencapai tujuan yang di harapkan, yaitu sebagai berikut:

1. Sebaiknya data atau citra wajah yang digunakan sebagai data *Training* dan *Testing* benar-benar menunjukkan posisi yang menampilkan semua komponen wajah (mata, hidung, mulut, alis), dan tidak menggunakan Aksesoris apapun di wajah.
2. Penulis berharap untuk penelitian selanjutnya agar mendapatkan hasil akurasi yang lebih baik dengan menggunakan dataset yang lain selain dataset yang di pakai dipenelitian ini dengan menggunakan metode yang sama yakni metode *Linear Discriminant Analysis* dan *C4.5*

DAFTAR PUSTAKA

- [1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, 1997, doi: 10.1109/34.598228.
- [2] M. Turk and A. Pentland, "Eigedces for Recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, 1991.
- [3] Rismayanti, "Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa Di Stt Harapan Medan," *Media Infotama*, vol. 12, no. 2, pp. 116–120, 2016.
- [4] F. Fandiansyah, J. Y. Sari, and I. P. Ningrum, "Pengenalan Wajah Menggunakan Metode Linear Discriminant Analysis dan k Nearest Neighbor," *J. Ultim.*, vol. 9, no. 1, pp. 1–9, 2017, doi: 10.31937/ti.v9i1.557.
- [5] A. Budi, S. Suma'inna, and H. Maulana, "Pengenalan Citra Wajah Sebagai Identifier Menggunakan Metode Principal Component Analysis (PCA)," *J. Tek. Inform.*, vol. 9, no. 2, pp. 166–175, 2018, doi: 10.15408/jti.v9i2.5608.
- [6] A. B. W.P and D. L. Pandiangan, "Pengenalan Pola Wajah untuk Simulasi Presensi Mahasiswa," *SMARTICS Journal*, vol. 2, no. 1. pp. 15–20, 2016.
- [7] T. I. Panggabean, "Pendeteksian dan Pengenalan Wajah Manusia untuk Peningkatan Kinerja Kamera Pengaman," 2018.
- [8] R. A. S, "Pengenalan Wajah Menggunakan Metode Fisherface," vol. 7, no. 1, pp. 50–61, 2007.
- [9] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014.

