

BAB III

METODOLOGI PENELITIAN

3. Metode Penelitian

1.1. Penelitian ini menggunakan jenis penelitian eksperimen, dengan tahapan penelitian sebagai berikut :

1. Analisis data Pengumpulan Data citra biji coklat
Tahap ini dilakukan sebagai awal dari suatu penelitian. Studi literature dilakukan pada literature – literature yang sesuai dengan penelitian ini.
2. Pengenalan data citra biji coklat
Pengenalan biji coklat didasari pada penyesuaian dengan citra yang ada di dalam database.
3. Rekayasa Citra biji coklat
Tahap ini dilakukan untuk mencari kemiripan atau *similarity*
4. Penentuan Kemiripan citra biji coklat
menghitung kemiripan (*similarity*) antara cintra biji coklat dengan citra biji coklat yang ada dalam database
5. Eksperimen dan pengujian metode
Tahap ini akan menjelaskan penelitian dan Teknik pengujian yang akan digunakan.
6. Evaluasi dan validasi hasil
Tahap ini akan menyimpulkan dari hasil evaluasi dari eksperimen yang dilakukan dari dataset yang telah dikumpulkan.

3.2 Analisis data dan Pengumpulan data citra biji coklat

Pada tahap ini akan dijelaskan tentang sumber data yang digunakan dan tahap *pra-processing* data.

1. Data
Sumber data yang digunakan dalam penelitian ini berupa 522 data
2. Tahap *Pra-Processing*
Sebelum data siap digunakan, data tersebut harus melewati tahap *pra-processing*. *Pra- processing* merupakan tahap untuk menyiapkan data mentah dengan tujuan agar data tersebut siap dipelajari. Proses ini dilakukan menggunakan bantuan dari Microsoft Excel.
 - a. *Data Selection*

Tahap pertama yang harus dilakukan adalah *data selection* yaitu dengan memilih atribut – atribut yang diperlukan serta membuang atribut yang tidak dibutuhkan. Adapun atribut yang terpilih untuk *data training* dapat dilihat pada tabel 3.1.

Tabel 3.1 Atribut Yang Digunakan Dalam Data Training

Atribut	Keterangan
Kadar air	Kadar air yang terkandung didalam biji coklat
Kotoran	Kotoran yang berada dalam biji coklat yang dempet, pecah plasenta dan serpihan buah.
Isi biji/gram	Ini adalah isi biji coklat per 100 gram
Jamur	Tingkat jamur yang terpapar pada biji coklat
Ukuran Biji	Ukuran biji dari biji coklat
berat	Berat ukuran setiap biji coklat

Atribut hasil merupakan atribut baru yang dibuat sebagai variabel keputusan. Atribut hasil dibuat berdasarkan study literatur menurut (SNI), selanjutnya akan diklasifikasi menjadi 3 keputusan yaitu Penggunaan bagus (premium), dan tidak bagus (Asalan). Atribut yang terpilih untuk data biji coklat yang akan digunakan sebagai *data testing* dapat dilihat pada tabel 3.2.

Tabel 3.2 Atribut Yang Digunakan Dalam *Data Testing*

Atribut	Keterangan
Kadar air	Kadar air yang terkandung didalam biji coklat
Kotoran	Kotoran yang berada dalam biji coklat yang dempet, pecah plasenta dan serpihan buah.
Isi biji/gram	Ini adalah isi biji coklat per 100 gram
Jamur	Tingkat jamur yang terpapar pada biji coklat
Ukuran biji	Ukuran biji dari biji coklat
berat	Berat ukuran setiap biji coklat

b. *Data Training*

Data Training adalah data yang digunakan untuk membentuk sebuah model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas

data baru yang belum pernah ada. Dalam hal ini digunakan sebuah pembagian data untuk data training sebanyak 522 data biji coklat.

c. *Data Testing*

Data Testing adalah sebuah data yang digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Karena itu, data yang ada pada testing tidak sama dengan data *training*. dan data testing menggunakan data biji coklat yang belum diketahui kualitasnya.

3.2.1. Tahapan Algoritma Naive Bayes

Dalam penelitian ini akan diimplementasikan metode *Naive Bayes* untuk menentukan prediksi kualitas biji coklat. Pada tahapan ini digunakan beberapa tahapan yang akan dilakukan dalam perhitungan *naive bayes* pada penelitian ini.

a. Penentuan atribut yang akan digunakan

Pada penelitian ini digunakan kadar air, kotoran, isi biji/gram, jamur, ukuran biji, dan berat. Berikut ini adalah rincian dari atribut dan nilai atribut yang digunakan dalam perhitungan Algoritma *Naive Bayes* dapat dilihat pada tabel 3.3.

Tabel 3.3 Atribut yang digunakan

Atribut	Nilai Atribut
Kadar air	<ul style="list-style-type: none">• 6 - 7 %• 7 – 8 %• 7,5%• 8 – 9 %• 10 % +
Kotoran	<ul style="list-style-type: none">• 0 %• 2 %• 2,5 %• 3 – 4 %• 5 %
Isi biji/gram	<ul style="list-style-type: none">• 85• 85 – 100• 101 – 110• 111 – 120• 120 +

Jamur	<ul style="list-style-type: none"> • 1 – 2 % • - • 4 % • 4 %+ • 5 – 6 %+
Ukuran biji	<ul style="list-style-type: none"> • seragam • tidak seragam
Berat	<ul style="list-style-type: none"> • Lebih dari 1 gram • Kurang dari 1 gram

b. Penentuan Data Set

Data set adalah data yang di gunakan untuk menggali informasi yang di gunakan sebagai data training. Data set dapat dilihat pada tabel 3.4 dan selengkapnya dapat dilihat pada lampiran data set.

Tabel 3.4 Data Set

Browse	Structure	SQL	Search	Insert	Export	Import	Privileges	Operations	Tracking	Triggers		
<input type="checkbox"/>				1	1	seragam	7,5	4	15	10	1 gram	premium
<input type="checkbox"/>				2	2	tidak seragam	8	2	11	14	kurang 1 gram	premium
<input type="checkbox"/>				3	3	seragam	6	3	44	9	1 gram	premium
<input type="checkbox"/>				4	4	seragam	7,5	4	20	8	1 gram	premium
<input type="checkbox"/>				5		seragam	7,5	6	16	12	1 gram	premium
<input type="checkbox"/>				6		tidak seragam	6	2	29	5	kurang 1 gram	premium
<input type="checkbox"/>				7		seragam	7	7	13	13	kurang 1 gram	premium
<input type="checkbox"/>				8		seragam	8	8	49	5	kurang 1 gram	premium
<input type="checkbox"/>				9		seragam	8	5	38	15	kurang 1 gram	premium
<input type="checkbox"/>				10		seragam	10	4	19	10	kurang 1 gram	asalan

3.3 Pengenalan data citra biji coklat

Dalam penelitian ini Pengenalan data citra biji coklat menggunakan citra yang ada di dalam database. Untuk itu dalam penelitian ini akan menggunakan teknik Matching Citra JPEG Berbasis Content Untuk CBIR dan menggunakan indexing untuk metode CBR dengan rumus naïve bayes.

3.4 Penentuan Kemiripan citra biji coklat

Dalam penelitian ini, citra yang dipakai adalah citra dengan format JPEG (citra didapatkan). Dalam citra JPEG, citra dibagi-bagi kedalam blok-blok (matriks 8x8) Perindexan (indexing) citra dilakukan dengan menggabungkan N blok kedalam suatu gabungan blok dan setiap blok bertindak sebagai satuan ruang dalam gabungan blok kunci indeks (indexing key) akan mempunyai 64 element dan setiap element mempunyai 1 lokasi. Pengindeksian kunci dapat dihitung dengan menggunakan formula berikut :

$$h_i = \frac{\sum_{i=1}^N (DCT_i)^2}{N}$$

dimana h_i adalah element ke I dari vector dan indexing key dapat dijelaskan sebagai berikut : $H_1 = \{h_0, h_1, \dots, h_{63}\}$ sedangkan untuk menghitung Euclidean distance dapat dihitung dengan formula :

$$d(h_q, h_k) = \left(\sum_{i=1}^N \frac{|h_i^q - h_i^k|}{64} \right)$$

$D(h_q, h_k)$ adalah jarak antara citra referensi dan citra dalam database. Beberapa aplikasi CBIR dengan menggunakan histogram warna, layout warna, tekstur local, koefisien wavelet, secara umum algoritma matching dapat dijelaskan sebagai berikut, misalnya terdapat dua histogram I dan M yang mana tiap histogram mempunyai storage (bin) sebanyak n, maka perpotongan dari kedua histogram dapat di definisikan sebagai berikut.

Setelah tahapan generating citra biji coklat berdasarkan indexing (CBIR) dan matching citra JPEG berbasis konten, maka kemudian untuk mencari kemiripan atau *similarity* peneliti akan menggunakan perhitungan matching citra JPEG berbasis konten (CBIR).

Secara umum algoritma matching dapat dijelaskan sebagai berikut, misalnya terdapat dua histogram I dan M yang mana tiap histogram mempunyai storage (bin) sebanyak n, maka perpotongan dari kedua histogram dapat didefinisikan sebagai berikut:

$$H(I, M) = \frac{\sum_{j=1}^n \min(I_j, M_j)}{\sum_{j=1}^n M_j}$$

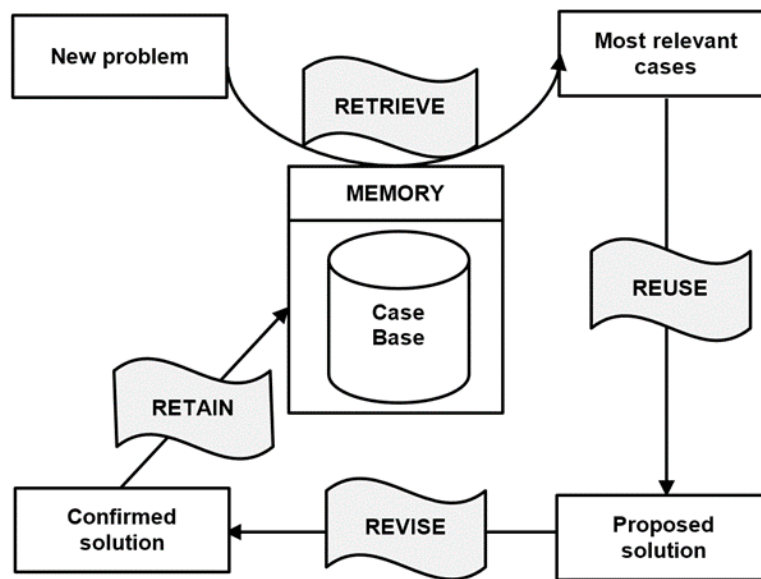
Dikarenakan kekurangan dan kelebihan image retrieval berbasis teks dan konten, maka beberapa peneliti telah mencoba menggabungkan keduanya untuk pencarian citra. Terdapat banyak pendekatan telah digunakan untuk menggabungkan teks atau keyword dan teknik berbasis konten untuk pengambilan gambar.

diperkenalkan oleh para penulis (Besançon et al 2005; Jones et al 2005)

3.5 Experiment dan pengujian metode

a. Metode CBR

Dalam melakukan experiment disini menggunakan 2 metode yang pertama metode CBR dan CBIR. Pertama disini akan membahas untuk metode CBR dimana CBR dengan membandingkan kasus baru dengan kasus lama. Jika kasus baru tersebut mempunyai kemiripan dengan kasus lama maka CBR akan memberikan solusi kasus lama untuk kasus baru tersebut. Kurang lebih jalan prosesnya seperti di gambar di bawah ini dengan mengacu pada 4 tahapan proses dari CBR



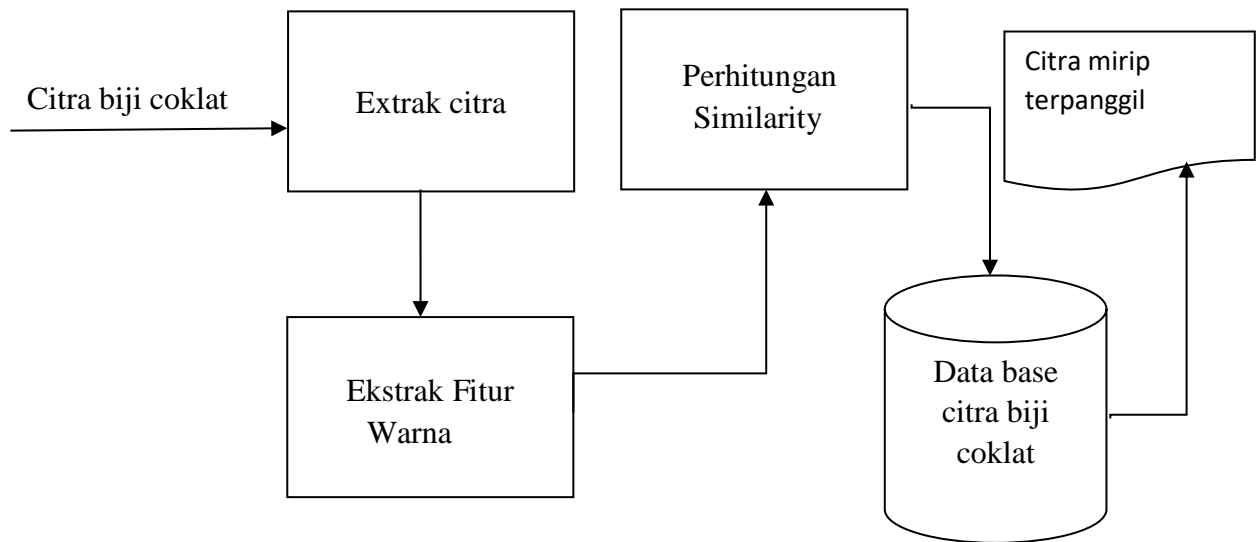
Gambar 3.1 Tahapan Proses Case Based Reasoning

b. Metode CBIR

Kedua untuk tahapan proses metode CBIR. CBIR adalah teknik atau metode yang menggunakan konten visual untuk mencari gambar dalam basis data berskala besar kita akan jabarkan proses metode CBIR yang akan di jalankan seperti gambar di bawah berikut.

Pertama citra coklat akan di ekstraksi dalam ekstraksi citra pertama akan mengekstraksi fitur dan ekstraksi warna setelah mendapatkan ekstraksi fitur dan warna maka akan di lanjutkan ke dalam perhitungan similarity dengan menghitung *Euclidean distance*-nya, Sementara itu untuk mengukur efektivitas dari pencarian gambar dengan menggunakan precision dan Recall, precision adalah jumlah citra yang mirip yang terpanggil dibagi dengan semua citra yang terpanggil (*retrieved*). Sedang Recall adalah jumlah citra yang terpanggil dibagi dengan jumlah citra yang dalam kategori.

Setelah dapat perhitungan maka hasil dari similarity image akan masuk ke dalam database untuk memanggil citra yang mirip dengan citra baru maka akan menghasilkan hasil dari citra baru.



Gamabar 3.2 Proses menggunakan metode CBIR