

BAB II

TINJAUAN PUSTAKA

1.1 Coklat

Coklat atau kakao adalah merupakan salah satu tanaman komoditas perkebunan di Indonesia yang memiliki peranan penting bagi perekonomian nasional, biji kakao mengandung senyawa polifenol yang berperan sebagai antioksidan, polifenol golongan flavonoid terutama katekin dan epikatekin adalah komponen utama dalam biji kakao [].Tanaman kakao menghasilkan biji kakao yang apabila dilihat dari warna belahan bijinya, kakao dibedakan menjadi dua tipe, yaitu warna putih termasuk ke dalam grup Criollo, yang disebut juga dengan *fine cocoa* atau kakao mulia. Sedangkan tanaman kedua, menghasilkan keping biji berwarna ungu, termasuk grup Forastero. Dari segi citarasa dan aroma, jenis pertama lebih baik tapi dari segi produktivitas, tanaman Forastero lebih tinggi. Biji kakao yang banyak diperdagangkan mayoritas adalah jenis kedua atau sering disebut biji kakao lindak atau *bulk cocoa*.

1.2 Citra digital

Citra digital adalah citra yang dapat diolah oleh komputer. Istilah citra digital sangat populer pada masa sekarang. Banyak peralatan elektronik, misalnya *scanner*, kamera digital, mikroskop digital, dan *fingerprint reader* (pembaca sidik jari), yang menghasilkan citra digital juga sangat populer digunakan oleh pengguna untuk mengolah foto atau untuk berbagai keperluan lain. Sebagai contoh, *Adobe Photoshop* dan *GIMP (GNU Image Manipulation Program)* menyajikan berbagai fitur untuk memanipulasi citra digital [1]

Berdasarkan warna-warna penyusunnya, citra digital dapat dibagi menjadi tiga macam [2] yaitu:

1. Citra biner, yaitu citra yang hanya terdiri atas dua warna, yaitu hitam dan putih. Oleh karena itu, setiap pixel pada citra biner cukup direpresentasikan dengan 1 bit. Citra biner sering kali muncul sebagai hasil dari proses pengolahan seperti segmentasi, pengembangan, morfologi, ataupun dithering.
2. Citra grayscale, yaitu citra yang nilai pixel-nya merepresentasikan derajat keabuan atau intensitas warna putih. Nilai intensitas paling rendah merepresentasikan warna hitam dan nilai intensitas paling tinggi merepresentasikan warna putih. Pada umumnya citra grayscale memiliki

kedalaman pixel 8 bit (256 derajat keabuan), tetapi ada juga citra grayscale yang kedalaman pixel-nya bukan 8 bit, misalnya 16 bit untuk penggunaan yang memerlukan ketelitian tinggi .

Citra berwarna, yaitu citra yang nilai pixel-nya merepresentasikan warna tertentu. Banyaknya warna yang mungkin digunakan bergantung kepada kedalaman pixel citra yang bersangkutan. Citra berwarna direpresentasikan dalam beberapa kanal (channel) yang menyatakan komponen- komponen warna penyusunnya. Banyaknya kanal yang digunakan bergantung pada model warna yang digunakan pada citra tersebut.

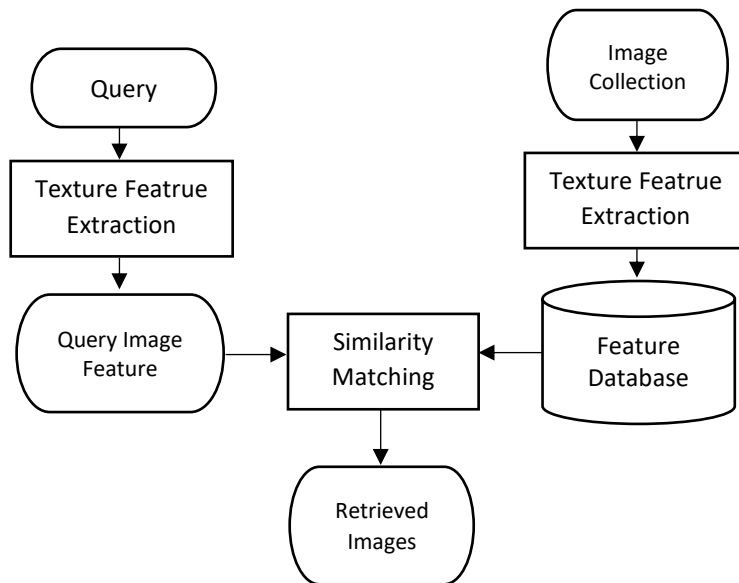
1.3 Content Base Image Retrival

Aplikasi *conten based image retrieval* telah secara luas digunakan di berbagai bidang kehidupan seperti bidang biomedis, kriminalitas, militer, commerce, budaya, Pendidikan, hiburan dan pertanian. Banyak proyek dan penelitian telah dikerjakan diantaranya adalah yang dikerjakan oleh IBM (<https://www.qbic.almaden.ibm.com>), Virage (www.virage.com), AltaVista (www.altavista.com), Yahoo (www.yahoo.com), and Google (www.google.com).

CBIR adalah Teknik atau metode yang menggunakan konten visual untuk mencari gambar dalam basis data bersekala besar. CBIR juga adalah Teknik yang terkenal banyak dipakai dalam pencarian citra berdasarkan konten visual seperti warna dan bentuk yang di ekstrak dalam bentuk vector. CBIR juga disebut sebagai aplikasi Teknik computer vision yang menangani masalah pencarian gambar dalam sekala besar.

Beberapa hal yang perlu di perhatikan dalam CBIR di antaranya adalah :

- Pemahaman user terhadap kebutuhan dan information seeking behavior.
- Identifikasi Teknik yang cocok bagaimana menerjemahkan content dari citra
- Ekstraksi fitur citra awal (raw image)
- Bagaimana menyediakan storage untuk databse citra yang biasanya cukup besar
- Matching citra query dan citra yang ada dalam database
- Efisiensi



Gambar 2.3 diagram alur *Content Based Image Retrieval*(CBIR)

Pada penelitian ini nantinya akan digunakan perhitungan *Euclidean Distance*. Menurut [3] Metode yang paling sering digunakan untuk menghitung kesamaan dua vektor salah satunya adalah *Euclidean Distance*. Adapun persamaan metode ini adalah sebagai berikut:

$$d(A, B) = \sqrt{\sum_{j=1}^n (H_j^A - H_j^B)^2}$$

Keterangan :

A : Vektor A

B : Vektor B

$d(A,B)$: Jarak *Euclidean* antara vektor A dan vektor B

n : Jumlah elemen vektor

H : Elemen vektor

Yang mana pada penelitian ini nantinya akan digunakan untuk menghitung akurasi kesamaan dua vektor antara database yang tersegmentasi

1.4 Segmentasi

Dalam pengolahan citra, terkadang kita menginginkan pengolahan hanya pada obyek tertentu. Oleh sebab itu, perlu dilakukan proses segmentasi citra yang bertujuan untuk memisahkan antara objek (*foreground*) dengan *background*.

Segmentasi citra adalah proses membagi citra digital menjadi beberapa daerah atau kelompok, dimana masing-masing daerah terdiri dari sekumpulan piksel. Segmentasi citra menyederhanakan dan merubah representasi citra kesesuatu yang lebih bermakna dan lebih mudah untuk dianalisis. Segmentasi citra digunakan untuk mencari obyek yang ingin dicari dan batas-batas bentuk objek seperti garis, kurva dalam citra [4].

Banyak algoritma segmentasi yang menggunakan parameter-parameter tertentu, segmentasi multi-resolusi juga memerlukan parameter-parameter seperti skala, warna, bentuk, kepadatan, dan kehalusan. Ukuran dan bentuk dari citra suatu objek dapat dipengaruhi oleh nilai parameter yang berbeda, perlu juga diketahui bahwa untuk menemukan ukuran dan bentuk objek yang sesuai untuk klasifikasi membutuhkan banyak waktu dan sangat subjektif

1.5 Matching kemiripan

untuk menghitung kemiripan (similarity) antara citra query dengan citra biji coklat yang ada dalam databse secara langsung di perlukan biaya yang cukup mahal dalam arti algoritma yang kompleks dan dan proses yang cukup lama. Untuk mengatasi masalah ini, maka dalam penelitian inidilakukan dalam tiga tahapan dalam matching citra biji coklat yaitu?

1. Pertama mencari kemiripan topologi dari citra query dengan citra dalam database yang merupakan filter
2. Kedua penggunaan informasi untuk memperbaiki calon citra yang akan terpanggil
3. Dan yang terakhir metode perehitungan matching dikerjakan untuk menentukan kemiripan (similarity) antara citra query dengan citra yang ada dalam data base.

Sementara itu untuk emngukur efektivitas dari pencarian gambar dengan menggunakan precision recall. Precision recall adalah jumlah citra yang mirip terpanggil dinagi dengan semua citra yang terpanggil (retrieved). Sedangkan recall adalah jumlah citra yang terpanggil dibagi jumlah citra yang dalam kaetagory (Cho,2004)

$$p = \frac{A}{Z} \quad r = \frac{b}{y}$$

Dmna p adalah precision dan r adalah recall sementara a adlah jumlah citra yang relevan atau mirip terpanggil. Z adalah jumlah semua citra yang terpanggil, dan y jumlah citra relevan.

1.6 JPEG (*Joint Photographic Export Group*)

JPEG merupakan sebuah standar untuk format citra yang diperkenalkan oleh *Joint Photographic Experts Group*. JPEG grup merupakan hasil kerja sama antara *International*

Standardization Organization (ISO) dan *International Electrotechnical Commission (IEC)*. JPEG pertama sekali dibuat pada tahun 1986 dan diperkenalkan pada tahun 1992. Pengembangan dan pembaharuan dilakukan seiring dengan berjalannya waktu sehingga pada tahun 1994 JPEG grup merilis versi terbaru JPEG. JPEG sering disebut sebagai standar yang cocok untuk algoritma kompresi *file* citra pada komputer dan menjadi salah satu format *file* citra yang didukung oleh *World Wide Web*. JPEG banyak digunakan untuk menyimpan gambar-gambar dengan ukuran *file* lebih kecil.[5]

1.7 Case Base Reasoning

Cara kerja CBR adalah 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. Jika tidak ada yang cocok, CBR akan melakukan proses adaptasi, yakni memperbaiki pengetahuan lama agar sesuai untuk menyelesaikan kasus baru. Kemudian pengetahuan baru akan disimpan sebagai salah satu basis kasus]. CBR menggunakan pendekatan kecerdasan buatan yang menitikberatkan pemecahan masalah berdasarkan pengetahuan dari kasus-kasus sebelumnya.

Definisi case base reasoning menurut Riesbaeck dan Schank (1989) adalah sebagai berikut :
“*sebuah penalaran berbasis kasus memecahkan masalah dengan menggunakan atau mengadaptasi solusi kasus lama*”

1.8 Kelebihan dan kekurangan case base reasoning

Penggunaan case base reasoning memiliki kelebihan dan kekurangan. Kelebihan dan kekurangan case base reasoning, adalah

- a. Pada case base reasoning tidak memerlukan pemahaman bagaimana menyelesaikan masalah sehingga mengurangi dampak penambahan informasi pengetahuan
- b. Pada case base reasoning pengetahuan di peroleh dengan cara mengumpulkan kejadian-kejadian yang telah terjadi yang tidak memerlukan suatu model eksplisit
- c. Pada case base reasoning memiliki kemampuan untuk belajar, cara yang dilakukan yaitu dengan menambahkan kasus baru seiring berjalannya waktu berjalan dan tanpa perlu menambahkan aturan baru atau mengubah yang sudah ada
- d. Kemampuan untuk mendukung justifikasi dengan menawarkan kasus lampau diutamakan

Sedangkan untuk kekurangan dari case base reasoning adalah :

- a. Memerlukan perhitungan yang kompleks untuk melakukan pengukuran kesamaan antar dua objek agar dapat penyelesaian masalah secara keseluruhan
- b. Seringkali menghadapi kesulitan dalam menentukan fitur-fitur dari suatu kasus agar bisa dibandingkan
- c. Semakin banyak kasus tersimpan dalam data base, maka dalam proses pencarian kasus yang paling mirip dengan kasus baru membutuhkan waktu yang cukup lama
- d. Solusi yang ditawarkan belum tentu yang terbaik, hal ini tergantung pada kualitas kasus-kasus sebelumnya yang telah tersimpan dalam database. Oleh karena itu maka tahapan revise menjadi hal yang penting untuk dilakukan sehingga mengurangi tingkat kesalahan

1.9 Tahapan case base reasoning

Ada 4 tahapan proses yang ada dalam case base reasoning yang disebut dengan siklus case base reasoning, yaitu

- a. Retrieve (penelusuran). Mendapatkan kasus-kasus yang mirip, pada tahap ini diawali dengan mengenali masalah dan diakhiri Ketika kasus yang akan dicari telah ditemui solusi yang mirip dengan kasus yang ada, tahapan yang akan dilakukan dalam proses penelusuran ini adalah:
 - Mengidentifikasi masalah
 - Pencocokan masalah
 - Penyeleksian masalah

Tahapan ini akan menentukan tingkat kemiripan (similarity) antar kasus yang dapat dicari dalam daftar dari kasus-kasus yang similar

- b. Reuse, menggunakan Kembali kasus-kasus yang ada dan dicoba untuk menyelesaikan suatu masalah sekarang focus reuse suatu kasus dalam kasus baru dipengaruhi oleh dua aspek, yaitu
 - Perbedaan antar kasus yang ada dengan kasus yang baru
 - Bagian mana dari penelusuran dari kasus yang dapat digunakan pada kasus baru
 Ada 2 (dua) cara yang dapat dilakukan dalam proses ini, adalah
 - Menggunakan Kembali solusi dari kasus yang telah ada (transformatial reuse)
 - Menggunakan Kembali metode kasus yang ada untuk membuat solusi (derivational reuse)

- c. Revise, merubah dan mengadopsi solusi yang ditawarkan jika perlu, ada dua tugas utama yang dimiliki pada tahap ini adalah mengevaluasi suatu solusi dan memperbaiki

kesalahan. Evaluasi suatu solusi adalah bagaimana hasil yang didapatkan setelah membandingkan solusi dengan keadaan yang sebenarnya. Perbaikan suatu kasus ini merupakan pengenalan kesalahan dari solusi yang di buat dan mengambil atau membuat penjelasan tentang kesalahan tersebut.

- d. Retrain, memakai solusi baru sebagai bagian dari kasus baru, kemudian kasus baru di update ke dalam basis kasus. tahapan ini terjadi suatu proses penggabungan dari solusi kasus yang baru dan benarke knowledge yang telah ada.

1.10 Representasi kasus

Ada beberapa pendekatan dalam representasi kasus yaitu :

- a. Konten dari sebuah kasus

Dalam representasi kasus oleh Janet Kolodner pada tahun 1991. Berfokus dalam peran kasus-kasus untuk membantu pengambilan keputusan pada konten dari kasus memegang peranan sangat penting

- b. Representasi fitur vector

System Protos (Porter, tahun 1990) menggunakan pendekatan fitur vector untuk domain yang lemah atau yang terlalu teoritis intractable theory

- c. Representasi berdasarkan frame

Representasi berdasarkan frame telah di rumuskan dengan deskripsi logika.

- d. Representasi objek oriented

Representasi objek oriented kasus memiliki sifat kemiripan (similarity) dengan representasi frame tapi memiliki asal yang berbeda

- e. Representasi tekstual

Representasi kasus tekstual seperti yang dideskripsikan oleh Lenz dan Buckhard (1996) mengurai teks yang ada didalam kasus menjadi entitas informasi. Entitas informasi dapat berupa kata atau frase yang terdapat dalam teks yang relevan untuk menentukan kemampuan pengguna Kembali dari bagian dicaptur dalam kasus

- f. Representasi kasus hirarki

Ide dasar pendekatan representasi kasus hierarki adalah untuk merepresentasikan sebuah kasus pada detail banyak level (multi level) memungkinkan pengguna untuk menggunakan banyak kata,

1.11 Algoritma naïve bayes

Algoritma *Naive Bayes* berbasiskan perhitungan probabilistik dengan asumsi bahwa setiap fitur yang digunakan saling lepas [].[6] Menyatakan *Naive Bayes* merupakan metode

klasifikasi teks yang paling populer digunakan. Algoritma ini memiliki kelebihan dari sisi kecepatan pembelajaran dan toleransinya terhadap nilai yang hilang dari fitur. Untuk menangani data numerik, algoritma ini menggunakan *probability density function*, artinya data dianggap mengikuti distribusi normal untuk kemudian dihitung nilai rata-rata dan simpangan bakunya.

Untuk merepresentasikan sebuah kelas, terdapat karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi yang berguna untuk menjelaskan bahwa peluang masuknya sampel karakteristik tertentu kedalam kelas *posterior*. Peluang munculnya suatu kelas (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik sampel secara global disebut juga *evidence*. Nilai *evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai *posterior* tersebut dibandingkan dengan nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel .

Klasifikasi Naive Bayes diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Persamaan dari teorema Bayes adalah sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Dengan :

X = Data dengan kelas yang belum diketahui;

H = Hipotesis data X merupakan suatu label kelas tertentu;

$P(H|X)$ = Probabilistik hipotesis H berdasarkan kondisi X (*posteriori probability*);

$P(H)$: Probabilistik hipotesis H (*prior probability*);

$P(X|H)$ = Probabilitas X berdasarkan kondisi pada hipotesis H;

$P(X)$ = Probabilistik X;

Untuk menjelaskan teorema *naive bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema *bayes* tersebut akan disesuaikan sebagai berikut:

$$P(C|E_1, \dots, E_n) = \frac{P(C)P(E_1, \dots, E_n|C)}{P(E_1, \dots, E_n)}$$

C = Sebuah kelas; F1 ... FN = Karakteristik Petunjuk.

$$\text{Posterior} = \frac{\text{Prior} \times \text{likelihood}}{\text{evidence}}$$

Penjabaran lebih lanjut sebagai Berikut :

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C) P(F_1, \dots, F_n | C) \\ &= P(C) P(F_1|C) \\ &= P(C) P(F_1|C) \\ &= P(C)P(F_1|C)P(F_2|C,F_1)P(F_3|C,F_1,F_2),P(F_4, \dots, F_n|C,F_1,F_2) \\ &= P(C)P(F_1|C)P(F_2|C,F_1)P(F_3|C,F_1,F_2), \dots, (F_n|C,F_1,F_2, \dots, F_{n-1}) \end{aligned}$$

Adapun langkah – langkah pada *naïve bayes* yang akan dilakakukan menggunakan rumus di atas sebagai berikut [7] :

1. Penentuan atribut yang akan digunakan .
2. Penentuan data set.
3. Lakukan perhitungan jumlah kasus pada atribut .
4. Tentukan kasus baru yang akan di klasifikasikan menggunakan metode naive bayes.
5. Klasifikasikan kasus baru berdasarkan kasus yang sama dengan kasus yang lama.
6. Kalikan semua kelas variable untuk mendapatkan nilai dari masing masing kelas.
7. Bandingkan hasil perkalian dari masing-masing kelas maka akan di ketahui berdasarkan hasil nilai terbesar dari masing-masing kelas.

1.12 Study literatur

Melakukan studi Pustaka dengan mengkaji jurnal-jurnal ilmiah dan berbagai sumber di internet mengenai penelitian yang berkaitan sebagai acuan untuk melakukan penelitian. Adapun referensi dalam penelitian ini sebagai berikut:

- I. Penelitian oleh [17] menggunakan metode naïve bayes untuk klasifikasi data dengan tingkat akurasi yang di dapat sebesar 86 %.
- II. Penelitian oleh [18]klasifikasi kualitas biji coklat menggunakan Support Vektor Machine akurasi yang didapat sebesar 82%.
- III. Penelitian oleh [14] system pakar diagnose penyakit tanaman kakao menggunakan metode *Case Base Reasoning* (CBR) pada kelompok tani gapoktan desa mekarti

jaya

- IV. Penelitian yang menggunakan metode *Region Growing* pernah dilakukan [19]. Jumlah data citra yang digunakan sebanyak 12 data dengan tingkat akurasi yang didapat sebesar 83,33%.
- V. Penelitian tentang *Region Growing* lainnya pernah dilakukan oleh [20]. Penelitian ini menggunakan metode *region growing* dengan parameter nilai tekstur yang didapat dari fitur kontras, homogenitas, entropi, dan energi pada metode *gray level co-occurrence matrix* (GLCM). Mendapatkan tingkat akurasi terbaik sebesar 84,7%.