

## BAB II

### KAJIAN PUSTAKA DAN DASAR TEORI

Dalam bab ini akan diuraikan beberapa kajian pustaka dan pembahasan teori dasar berkaitan dengan sistem identifikasi uang kertas rupiah menggunakan metode *modified k-nearest neighbour*. Kajian pustaka juga membahas beberapa penelitian yang sudah dilakukan sebelumnya.

#### 2.1 Kajian Pustaka

Pada penelitian oleh Meita Setiawan dan Albert Justinus, pengenalan uang kertas sudah pernah di aplikasikan menggunakan metode *template matching* dan objeknya adalah uang kertas dollar amerika. Pada penelitiannya berhasil di dapat akurasi tertinggi 100% dan pengenalan terendah adalah 78.6%. Tetapi pada penelitian pengenalan nominal uang menggunakan metode *template matching* ini, sistem tidak dapat mengenali uang jika kondisi uang dalam kondisi tidak baik, terdapat coretan, dan lusuh. Jadi jika ada uang dalam kondisi seperti diatas maka tingkat akurasi juga akan turun [SAN-07].

Pada skripsi ini identifikasi nominal uang akan dilakukan menggunakan metode *modified k-nearest neighbour*. Pada penelitian sebelumnya metode *modified k-nearest neighbour* sudah pernah diterapkan untuk mengidentifikasi penyakit tanaman kedelai [ZDD-14]. Pada penelitian identifikasi penyakit tanaman kedelai ini, dilakukan pengujian dengan 3 data uji, dan setiap pengujian dilakukan dengan nilai k yang berbeda-beda. Jadi pada penelitian ini dapat diketahui pengaruh nilai k terhadap akurasi algoritma *modified k nearest neighbour*. Dan didapatkan akurasi rata-rata tertinggi 92,74% terdapat pada saat k bernilai 3. Pada penelitian ini semakin tinggi nilai k maka hasil akurasi juga akan semakin kecil.

Pada penelitian selanjutnya yang berjudul “A Modification on K-Nearest Neighbour Classifier” membahas tentang modifikasi pada algoritma KNN dengan menambahkan perhitungan nilai validasi, sehingga dapat menghasilkan akurasi yang lebih tinggi dari algoritma KNN biasa. Pada penelitian yang dilakukan oleh Hamid Parvin, Hoseinali Alizadeh, dan Behrouz Minati ini menghasilkan

kesimpulan bahwa tingkat akurasi algoritma MKNN lebih tinggi daripada tingkat akurasi KNN biasa. Seperti pada percobaannya pada dataset Wine metode KNN mempunyai tingkat akurasi 83,79% sedangkan metode MKNN 85,76% dan juga pada dataset Isodata metode mempunyai tingkat akurasi KNN 82,90% sedangkan metode MKNN 83,32%. Jadi berdasarkan pada penelitiannya dihasilkan bahwa metode MKNN lebih mempunyai tingkat akurasi yang lebih tinggi dibandingkan algoritma KNN biasa [PRV-08]. Jadi agar identifikasi nominal uang ini juga memiliki akurasi yang tinggi maka juga akan menggunakan metode *modified k nearest neighbour*.

## 2.2 Uang

Kehidupan manusia umumnya tidak pernah terlepas dari uang. Uang merupakan bagian yang demikian besar dalam kehidupan manusia sehari-hari. Uang dalam ilmu ekonomi tradisional didefinisikan sebagai setiap alat tukar yang dapat diterima secara umum. Alat tukar itu berupa benda apa saja yang dapat diterima oleh setiap orang di masyarakat dalam proses pertukaran barang dan jasa. Uang adalah segala sesuatu yang merupakan media pertukaran atau alat pembayaran yang diterima secara umum.

Tahapan-tahapan perkembangan uang adalah yang pertama tahap sebelum barter, pada tahap ini masyarakat belum mengenal pertukaran karena setiap orang masih berusaha untuk memenuhi kebutuhan hidupnya dengan usaha sendiri. Lalu tahap Barter, tahap perkembangan selanjutnya menghadapkan manusia kepada kenyataan bahwa apa yang diproduksi sendiri tidak cukup untuk memenuhi kebutuhannya. Untuk memperoleh barang-barang yang tidak dapat dihasilkan sendiri mereka mencari orang yang mau menukarkan barang yang dimilikinya dengan barang lain yang dibutuhkannya. Akibatnya barter, yaitu barang yang ditukar dengan barang.

Sejalan dengan perkembangan perekonomian, maka perkembangan tukar-menukar yang harus dilayani dengan uang logam juga berkembang, sedangkan jumlah logam mulia (emas dan perak) terbatas. Penggunaan uang logam juga sulit dilakukan untuk transaksi dalam jumlah besar (sulit dalam pengangkutan dan penyimpanan). Sehingga lahirlah uang kertas. Mula-mula uang kertas yang

beredar merupakan bukti-bukti pemilikan emas dan perak sebagai alat/perantara untuk melakukan transaksi. Dengan kata lain, uang kertas yang beredar pada saat itu merupakan uang yang dijamin 100% dengan emas atau perak yang disimpan di pandai emas atau perak dan sewaktu-waktu dapat ditukarkan penuh dengan jaminannya [YSN-07].

### 2.2.1 Jenis Uang Berdasarkan Bahan

Jenis uang jika dilihat dari bahan untuk membuatnya terdiri dari dua macam, yaitu uang logam dan uang kertas.

1) Uang logam adalah uang yang dibuat dari semacam logam tertentu dengan berat dan kadar tertentu pula. Uang yang terbuat dari logam pada umumnya memiliki nilai nominal kecil, yang dibuat dengan ciri-ciri khusus untuk menghindari pemalsuan. Uang logam di Indonesia pada saat ini terdiri atas uang yang bernilai nominal mulai dari Rp50,00; Rp100,00; Rp200,00; Rp500,00; dan Rp1.000,00

2) Uang kertas merupakan uang fiduciary (uang kepercayaan), karena semua masyarakat mau menerima uang tersebut sebagai alat pembayaran, walaupun nilai intrinsiknya jauh lebih kecil daripada nilai nominalnya. Jadi, dasar uang kertas adalah kepercayaan kepada pemerintah atau bank yang menjamin atas peredaran uang kertas tersebut. Di samping kepercayaan umum, terdapat alasan lain yang mendorong untuk menciptakan uang kertas sebagai alat pertukaran, yaitu:

- Uang logam tidak dapat digunakan untuk jumlah yang sangat besar, sedangkan uang kertas tidak ada kesulitan,
- Biaya untuk membuat uang logam jauh lebih mahal daripada untuk membuat uang kertas,
- Uang logam kurang praktis, sukar dibawa ke tempat yang jauh dalam jumlah besar.
- Uang kertas yang beredar di masyarakat saat ini mulai dari pecahan Rp1.000,00; Rp5.000,00; Rp10.000,00; Rp20.000,00; Rp50.000,00; dan Rp100.000,00.

Semua uang kertas ini dicetak oleh Perum Peruri (Perusahaan Umum Percetakan Uang Republik Indonesia) dan peredarannya diatur oleh Bank

Indonesia. Oleh karena itu, uang kertas tersebut dinamakan uang kertas bank [YSN-07].

### 2.3 Automatic Ticket-Vending Machine

Sistem identifikasi uang kertas dapat diaplikasikan pada suatu anjungan mandiri swalayan, atau disebut Automatic Ticket – Vending Machine (ATVM). ATVM ini berfungsi menggantikan transaksi pembelian tiket atau makanan secara manual (non operator) ke mesin otomatis. Alat ini dapat melayani konsumen secara cepat, efisien dan mampu mengenali alat pembayaran yang berlaku sesuai dengan harga yang tertera [WCK-07].

Identifikasi uang hanyalah satu bagian diluar fungsi komputasi matematis mesin. AVTM dapat diletakkan ditempat-tempat umum dan stasiun / terminal untuk melayani konsumen yang ingin bertransaksi. Bentuk fisik dari ATVM dapat dilihat pada gambar 2.1 dibawah ini.



**Gambar 2. 1. ATVM : (a) penjual tiket (b) atm setor tunai**

Gambar (a) merupakan alat penjualan tiket otomatis sedangkan gambar (b) adalah alat setor tunai pada atm-atm yang sudah ada di Indonesia. Alat-alat ini tidak hanya untuk pembelian tiket dan setor tunai bank saja, tetapi sudah banyak diaplikasikan untuk penjualan minuman, rokok, permen dan lain-lain. Dan alat-alat tersebut juga dapat menerima pembayaran seperti uang koin, kertas, dan kartu kredit.

Dalam mesin ATVM terdapat suatu alat validasi yang memiliki mekanisme menarik uang kertas yang dimasukkan kemudian mengidentifikasinya. Alat validasi inilah yang mempunyai sistem komputasi cerdas dalam memvalidasi

benda yang masuk dalam mesin tersebut. Dan dalam penelitian ini nantinya dapat digunakan sebagai acuan dalam proses memvalidasi identifikasi uang kertas.

#### 2.4 Sistem Visual

Sistem visual buatan atau vision system (computer vision) adalah suatu sistem yang mempunyai kemampuan untuk menganalisa obyek secara visual, setelah data obyek yang bersangkutan dimasukkan dalam bentuk citra (image) [BAL-82].

Secara umum tujuan dari sistem visual adalah untuk membuat model nyata dari sebuah citra. Citra yang dimaksud adalah citra digital hasil konversi suatu obyek menjadi citra melalui suatu sensor yang prosesnya disebut digitasi. Dengan demikian citra lain seperti foto, gambar cetak, gambar sketsa, dan lain-lain yang berada pada media cetak seperti kertas atau media lainnya baru dapat diproses setelah dikonversi ke dalam citra digital melalui proses digitasi [BAL-82].

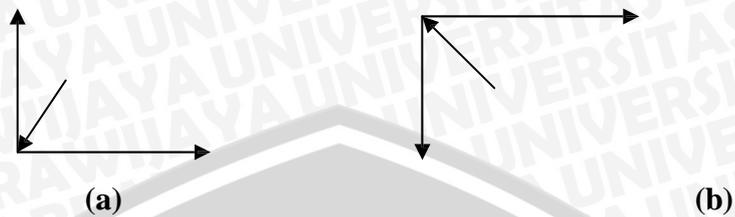
Sebuah sistem visual mempunyai kemampuan untuk memperbaiki informasi yang berguna dari sebuah pemandangan (scene) hasil proyeksi dua dimensi. Citra merupakan hasil proyeksi dua dimensi dari obyek atau benda tiga dimensi, sehingga informasi tidak bisa didapatkan begitu saja, melainkan harus diperbaiki karena sesungguhnya ada bagian informasi yang hilang setelah benda diproyeksikan ke dalam citra.

Dalam skripsi ini citra yang telah didapatkan akan diambil fitur yang dibutuhkan dan diolah menggunakan proses metode modified k nearest neighbour agar aplikasi ini dapat mengidentifikasi citra uang kertas.

#### 2.5 Citra Digital

Sebuah citra adalah kumpulan piksel-piksel yang disusun dalam larik dua dimensi. Piksel adalah sampel dari pemandangan yang mengandung intensitas citra yang dinyatakan dalam bilangan bulat. Indeks baris dan kolom (x,y) dari sebuah piksel dinyatakan dalam bilangan bulat. Piksel (0,0) terletak pada sudut kiri atas pada citra, indeks x bergerak ke kanan dan indeks y bergerak ke bawah.

Konvensi ini dipakai merujuk pada cara penulisan larik yang digunakan dalam pemrograman komputer [GZZ-08].



**Gambar 2. 2 Perbedaan letak titik origin pada kordinat grafik dan pada citra**  
**(a) koordinat pada grafik matematika (b) koordinat pada citra**

Ditinjau dari sudut pandang matematis, citra merupakan fungsi menerus (continue) dari intensitas cahaya pada bidang dwimatra. Sumber cahaya menerangi objek, objek memantulkan kembali sebagian dari berkas cahaya tersebut. Pantulan cahaya ini ditangkap oleh alat-alat optik, misalnya mata pada manusia, kamera, pemindai (scanner), dan sebagainya, sehingga bayangan objek yang disebut citra tersebut terekam.

Citra sebagai keluaran dari suatu sistem perekaman data dapat bersifat [MUR92]:

1. Optik berupa foto,
2. Analog berupa sinyal video seperti gambar pada monitor televisi,
3. Digital yang dapat langsung disimpan pada suatu pita magnetik.

Citra yang dimaksudkan di dalam keseluruhan skripsi ini adalah “citra diam” (still images). Citra diam adalah citra tunggal yang tidak bergerak. Gambar 2.3 adalah contoh 2 gambar citra diam berupa foto.



(a) Lena

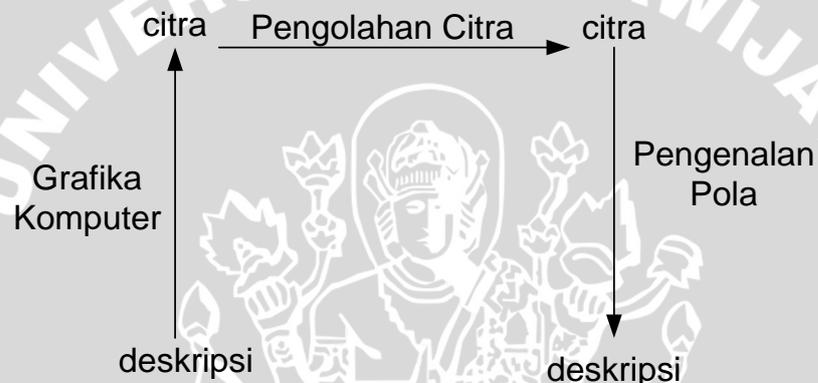
(b) Kapal

**Gambar 2. 3 Gambar Citra Lena dan cita kapal**

Di dalam bidang komputer, sebenarnya ada tiga bidang studi yang berkaitan dengan data citra, namun tujuan ketiganya berbeda, yaitu:

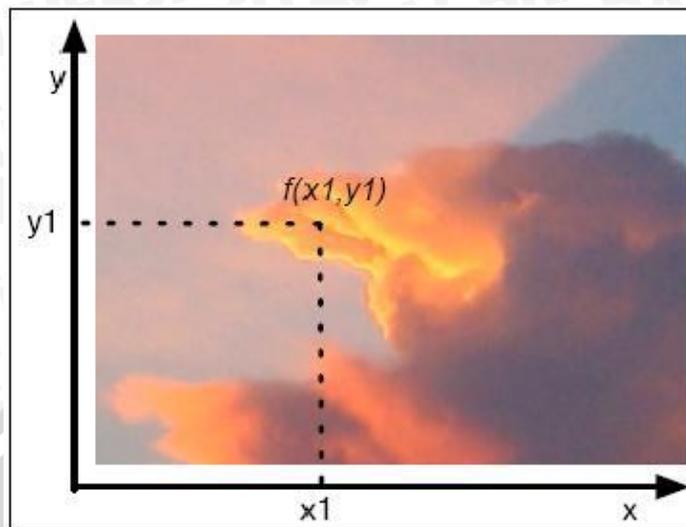
1. Grafika Komputer (computer graphics).
2. Pengolahan Citra (image processing).
3. Pengenalan Pola (pattern recognition/image interpretation).

Hubungan antara ketiga bidang (grafika komputer, pengolahan citra, pengenalan pola) ditunjukkan pada Gambar 2.4.

**Gambar 2. 4 Gambar Hubungan Ketiga Bidang (Grafika Komputer, Pengolahan Citra, Pengenalan Pola)**

## 2.6 Representasi Warna

Teknologi dasar untuk menciptakan dan menampilkan warna pada citra digital berdasarkan pada penelitian bahwa sebuah warna merupakan kombinasi dari tiga warna dasar, yaitu merah, hijau, dan biru (Red, Green, Blue - RGB). Contoh gambar citra digital dapat dilihat pada gambar 2.5 berikut.

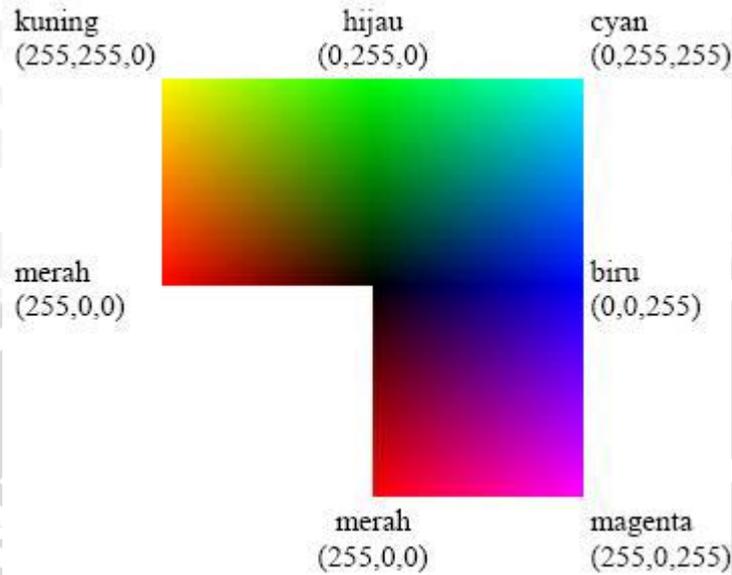


**Gambar 2. 5 Contoh Citra Digital**

Citra berwarna adalah citra yang lebih umum. Warna yang terlihat pada citra bitmap merupakan kombinasi dari tiga warna dasar, yaitu merah, hijau, dan biru. Setiap piksel disusun oleh tiga komponen warna : R (*red*), G (*green*), B (*blue*). Kombinasi dari tiga warna RGB tersebut menghasilkan warna yang khas untuk piksel yang bersangkutan. Pada citra 256 warna, setiap piksel panjangnya 8 bit, tetapi komponen warna RGB-nya disimpan di dalam tabel RGB yang disebut palet [MNR-04].

RGB adalah suatu model warna yang terdiri dari merah, hijau, dan biru, digabungkan dalam membentuk suatu susunan warna yang luas. Setiap warna dasar, misalnya merah, dapat diberi rentang-nilai. Untuk monitor komputer, nilai rentangnya paling kecil = 0 dan paling besar = 255. Pilihan skala 256 ini didasarkan pada cara mengungkap 8 digit bilangan biner yang digunakan oleh mesin komputer. Dengan cara ini, akan diperoleh warna campuran sebanyak  $256 \times 256 \times 256 = 1677726$  jenis warna. Sebuah jenis warna, dapat dibayangkan sebagai sebuah vektor di ruang 3 dimensi yang biasanya dipakai dalam matematika, koordinatnya dinyatakan dalam bentuk tiga bilangan, yaitu komponen-x, komponen-y dan komponen-z. Misalkan sebuah vektor dituliskan sebagai  $r = (x, y, z)$ . Untuk warna, komponen-komponen tersebut digantikan oleh komponen R(ed), G(reen), B(lue). Jadi, sebuah jenis warna dapat dituliskan sebagai berikut: warna = RGB(30, 75, 255). Putih = RGB (255,255,255), sedangkan untuk hitam=

RGB(0,0,0). Bentuk Representasi warna dari sebuah citra digital dapat dilihat pada Gambar 2.6.

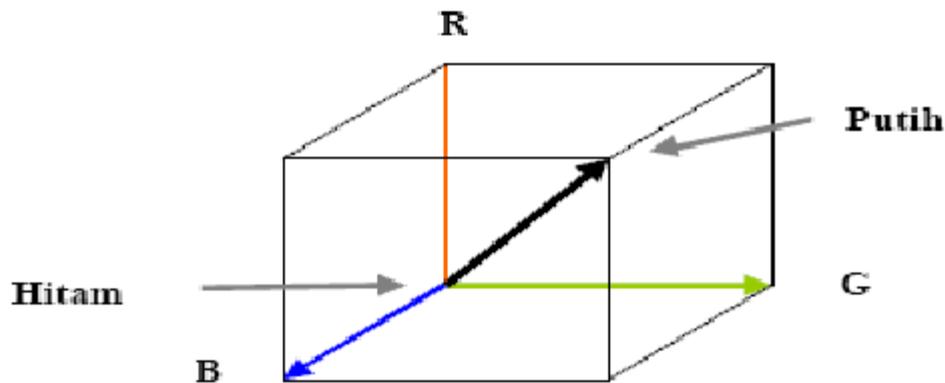


**Gambar 2. 6 Representasi warna RGB pada citra digital.**

Pada umumnya, warna dipisahkan menjadikan komponen-komponen, baik komponen warna ataupun komponen kecerahan, penyajian semacam ini disebut komponen warna. Pada komponen warna, tiap komponen dipisahkan dengan model-model tertentu, seperti model RGB, YUV dan YIQ [MNR-04].

### 2.6.1 RGB

Red Green Blue (RGB) merupakan warna dasar yang ada pada sistem komputer. Data video dapat dipisahkan menjadi komponen-komponen untuk masing-masing warna, yaitu merah (red), hijau (green) dan biru (blue). Warna tiap piksel ditentukan oleh kombinasi intensitas dari masing-masing komponen warna dan dipetakan dalam bentuk sebuah koordinat seperti terlihat pada gambar 2.7.



**Gambar 2. 7 Koordinat RGB**

Sebagai contoh, pada RGB 24 bit, masing-masing komponen warna dinyatakan dalam 8 bit atau 256 level. Misalnya, citra dengan 8 bit per piksel mempunyai 256 warna dan citra dengan 24 bit mempunyai 32768 warna, jadi setiap piksel dapat dinyatakan sebagai berikut :

1. Bit 0 sampai dengan 7 untuk warna merah.
2. Bit 7 sampai dengan 15 untuk warna hijau.
3. Bit 16 sampai dengan 24 untuk warna biru.

Kemungkinan kombinasi warna yang ada adalah  $= 256^3 + 256^2 + 256^1 = 16.843.008$ , dimana nilai 0 menyatakan warna hitam sedangkan nilai 16.843.008 menyatakan warna putih [MNR-04].

## 2.7 Pengenalan pola

Pengenalan pola (pattern recognition) dapat diartikan sebagai proses klasifikasi dari objek atau pola menjadi beberapa kategori atau kelas. Dan bertujuan untuk pengambilan keputusan [THD-92].

Pola adalah bentuk atau model yang dapat dipakai untuk membuat atau untuk menghasilkan suatu atau bagian dari sesuatu, khususnya jika sesuatu yang ditimbulkan mempunyai sejenis pola dasar yang dapat ditunjukkan atau terlihat, yang mana dapat dikatakan mempertunjukkan pola.

Ada beberapa definisi lain tentang pengenalan pola, di antaranya:

1. Penentuan suatu objek fisik atau kejadian ke dalam salah satu atau beberapa kategori.

2. Ilmu pengetahuan yang menitikberatkan pada deskripsi dan klasifikasi (pengenalan) dari suatu pengukuran.
3. Suatu pengenalan secara otomatis suatu bentuk, sifat, keadaan, kondisi, susunan tanpa keikutsertaan manusia secara aktif dalam proses pemutusan.

Berdasarkan beberapa definisi di atas, pengenalan pola dapat didefinisikan sebagai cabang kecerdasan yang menitik-beratkan pada metode pengklasifikasian objek ke dalam klas-klas tertentu untuk menyelesaikan masalah tertentu. Salah satu aplikasinya adalah pengenalan suara, klasifikasi teks dokumen dalam kategori (contoh. surat-E spam/bukan-spam), pengenalan tulisan tangan, pengenalan kode pos secara otomatis pada sampul surat, atau sistem pengenalan wajah manusia. Aplikasi ini kebanyakan menggunakan analisis citra bagi pengenalan pola yang berkenaan dengan citra digital sebagai input ke dalam sistem pengenalan pola.

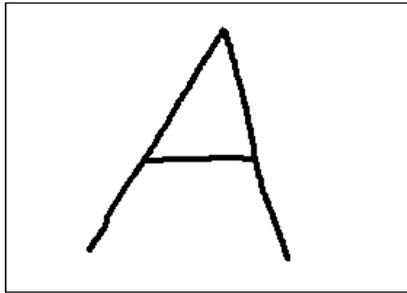
Pengenalan Pola mengelompokkan data numerik dan simbolik (termasuk citra) secara otomatis oleh mesin (dalam hal ini komputer). Tujuan pengelompokan adalah untuk mengenali suatu objek di dalam citra. Manusia bisa mengenali objek yang dilihatnya karena otak manusia telah belajar mengklasifikasi objek-objek di alam sehingga mampu membedakan suatu objek dengan objek lainnya. Kemampuan sistem visual manusia inilah yang dicoba ditiru oleh mesin. Komputer menerima masukan berupa citra objek yang akan diidentifikasi, memproses citra tersebut, dan memberikan keluaran berupa deskripsi objek di dalam citra [MNR-04]. Proses citra menjadi deskripsi dapat dilihat pada gambar 2.8 berikut.



**Gambar 2. 8 Pengenalan Pola**

Contoh pengenalan pola misalnya citra pada Gambar 2.9 adalah tulisan tangan yang digunakan sebagai data masukan untuk mengenali karakter 'A'.

Dengan menggunakan suatu algoritma pengenalan pola, diharapkan komputer dapat mengenali bahwa karakter tersebut adalah 'A'.



**Gambar 2. 9** Citra karakter 'A' yang digunakan sebagai masukan untuk pengenalan huruf.

Pada pengenalan pola dalam skripsi ini akan dilakukan untuk mengenali sebuah citra yang berbentuk foto. Yaitu untuk mengenali atau mengidentifikasi nominal uang. Citra yang digunakan adalah sebuah foto uang yang diperoleh dari hasil scanner dan disimpan dalam bentuk image jpg/png yang image tersebut nantinya akan diproses diambil fitur warnanya, dan dari fitur warna tersebut akan diolah lagi menggunakan metode modified k nearest neighbour.

## 2.8 Klasifikasi

Klasifikasi merupakan salah satu teknik data mining yang memiliki kemampuan untuk melakukan proses klasifikasi data. Klasifikasi bisa digunakan untuk menemukan model atau fungsi membedakan kelas data. Klasifikasi bertujuan untuk memprediksi kelas dari suatu objek yang labelnya tidak diketahui.

Klasifikasi merupakan suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan menggunakan hasilnya untuk memberikannya sejumlah aturan. Aturan-aturan tersebut digunakan pada data-data baru untuk diklasifikasi. Teknik ini menggunakan *supervised induction*, yang memanfaatkan kumpulan pengujian dan record yang terklasifikasi untuk menentukan kelas-kelas tambahan [KNW-07].

Klasifikasi merupakan proses untuk menyatakan suatu objek ke dalam salah satu kategori yang sudah ditentukan sebelumnya. Tujuan dari klasifikasi ini adalah

*record-record* yang sebelumnya belum termasuk dalam kategori dapat dinyatakan secara akurat. Tahapan klasifikasi terdiri dari [THD-92]:

1. Pembangunan model

Dalam tahapan ini dibuat model untuk menyelesaikan masalah klasifikasi data, model ini dibangun berdasarkan training set.

2. Penerapan Model

Pada tahap ini model sudah dibangun sebelumnya digunakan untuk menentukan atribut atau class dari sebuah data baru yang atribut atau classnya belum diketahui.

3. Evaluasi

Dalam tahapan ini hasil dari tahapan sebelumnya dievaluasi menggunakan parameter terukur untuk menentukan apakah model tersebut dapat diterima.

## 2.9 Min-Max Normalisasi

Metode normalisasi ini menghasilkan transformasi *linier* pada data asal. Bila  $minA$  dan  $maxA$  adalah nilai minimum dan maksimum dari sebuah atribut  $A$ , *Min-max Normalization* memetakan sebuah nilai  $v$  dari  $A$  menjadi  $v'$  dalam *range* nilai minimal dan maksimal yang baru,  $new\_minA$  dan  $new\_maxA$  [JYI-12]. Untuk menjaga agar setiap anggota  $v$  senantiasa terletak pada interval  $[0,1]$ , maka perlu dilakukan normalisasi. Rumus *Min-max Normalization* dapat dilihat pada persamaan (2-1).

$$V' = \frac{V - minA}{maxA - minA} \times (new_{maxA} - new_{minA}) + new_{minA} \quad \dots(2-1)$$

$V'$  = Nilai dari data hasil Min-Max Normalization

$V$  = Nilai dari data yang akan dinormalisasi

$minA$  = Nilai minimum dari suatu field data yang sama

$maxA$  = Nilai maksimum dari suatu field data yang sama

$new_{minA}$  = Nilai minimum terbaru yang diinginkan

$new_{maxA}$  = Nilai maksimum terbaru yang diinginkan

Keuntungan dari metode ini adalah keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses normalisasi. Tidak ada data bias yang dihasilkan oleh metode ini.

## 2.10 Algoritma K-Nearest Neighbour

Prinsip kerja *K-Nearest Neighbor* (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan  $k$  tetangga (*neighbor*) terdekatnya dalam data pelatihan [RWA-08].

Teknik ini termasuk dalam kelompok klasifikasi nonparametric. Di sini tidak memperhatikan distribusi dari data yang ingin di kelompokkan. Teknik ini sangat sederhana dan mudah diimplementasikan. Mirip dengan teknik klustering, yaitu dengan mengelompokkan suatu data baru berdasarkan jarak data baru itu ke beberapa data/tetangga (*neighbor*) terdekat [STB-07].

Tujuan algoritma KNN adalah mengklasifikasikan obyek baru berdasarkan atribut dan *training sample*. *Clasifier* tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik *query*, akan ditemukan sejumlah  $k$  obyek atau (titik training) yang paling dekat dengan titik *query*. Klasifikasi menggunakan voting terbanyak diantara klasifikasi dari  $k$  obyek. Algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru. Algoritma metode KNN sangatlah sederhana, bekerja berdasarkan jarak terpendek dari *query instance* ke *training sample* untuk menentukan KNN-nya.

Nilai  $k$  yang terbaik untuk algoritma ini tergantung pada data. Secara umum, nilai  $k$  yang tinggi akan mengurangi efek *noise* pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi semakin kabur. Nilai  $k$  yang bagus dapat dipilih dengan optimasi parameter. Kasus khusus dimana klasifikasi diprediksikan berdasarkan training data yang paling dekat (dengan kata lain,  $k=1$ ) disebut algoritma Nearest Neighbor [YYE-10].

Kelebihan KNN (*K-Nearest Neighbor*) [YYE-10]:

1. Tangguh terhadap *training* data yang memiliki banyak *noise*.
2. Efektif apabila *training* datanya besar.

Kelemahan KNN (*K-Nearest Neighbor*) [YYE-10]:

1. KNN perlu menentukan nilai dari parameter  $k$  (jumlah dari tetangga terdekat).

2. *Training* berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan.
3. Atribut mana yang harus digunakan untuk mendapatkan hasil terbaik.
4. Biaya komputasi cukup tinggi karena diperlukan perhitungan jarak dari tiap *query instance* pada keseluruhan *training sample*.

Algoritma KNN (*K-Nearest Neighbor*) :

1. Tentukan parameter *k*
2. Hitung jarak antara data yang akan dievaluasi dengan semua pelatihan
3. Urutkan jarak yang terbentuk (urut naik)
4. Tentukan jarak terdekat sampai urutan *k*
5. Pasangkan kelas yang bersesuaian
6. Cari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi

Rumus Jarak Eucliden:

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \dots(2-1)$$

Keterangan:

$x_1$  = Sampel Data

$x_2$  = Data Uji / Testing

$i$  = Variabel Data

$d$  = Jarak

$p$  = Dimensi Data

### 2.11 Metode Modified K-Nearest Neighbour

Modified K-Nearest Neighbor(MKNN) adalah menempatkan label kelas data sesuai dengan *k* divalidasi poin data yang sudah ditetapkan dengan perhitungan KNN. Dalam proses algoritma mknn, terdapat beberapa tambahan proses dibandingkan dengan knn, yaitu, validasi data training dan weight voting. Tujuan utama dari metode ini adalah memasukkan label kelas dari data berdasarkan data poin pada data latih yang sudah divalidasi dengan nilai *k*. Dengan kata lain, pertama dilakukan perhitungan validitas data pada semua data

di data latih. Selanjutnya, dilakukan perhitungan untuk mencari Weight Voting pada semua data uji menggunakan validitas data [PHD-10].

### 2.11.1 Validitas data training

Dalam metode MKNN setiap data training harus divalidasi terlebih dahulu. Validitas setiap data tergantung pada setiap tetangganya. Setelah dihitung validitas tiap data maka nilai validitas tersebut akan digunakan sebagai informasi lebih mengenai data training. Validasi digunakan untuk menghitung jumlah titik dengan label yang sama untuk data training.

Untuk menghitung validitas dari data pada data training, tetangga terdekatnya perlu dipertimbangkan. Diantara tetangga terdekat dengan data, validitas digunakan untuk menghitung jumlah titik dengan label yang sama untuk data tersebut.

Persamaan (2-2) digunakan untuk menghitung validitas dari setiap data adalah sebagai berikut:

$$\text{Validitas}(x) = \frac{1}{k} \sum_{i=1}^k S(\text{label}(x), (\text{label}(N_i(x)))) \quad \dots(2-2)$$

Dimana:

K = jumlah titik terdekat

Label (x) = kelas x

Label  $N_i(x)$  = label kelas titik terdekat x

Fungsi S digunakan untuk menghitung kesamaan antara titik x dan data ke-I tetangga terdekat. Persamaan untuk mendefinisikan fungsi S terdapat dalam persamaan,

$$S(a, b) = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases} \quad \dots(2-3)$$

Dimana:

a = kelas a pada data training

b = kelas selain a pada data training

Melalui persamaan (2-3) ditunjukkan bahwa a dan b adalah label kelas kategori suatu data latih. S akan bernilai 1, jika label kategori a sama dengan label kategori b. S akan bernilai 0, jika kategori a tidak sama dengan label kategori b [PHD-10].

### 2.11.2 Weight Voting

Weight voting adalah salah satu variasi metode KNN yang digunakan k tetangga terdekat dan hasil perhitungan dari jarak dari masing-masing data. Dalam metode KNN, digunakan rumus *Weight* untuk masing-masing tetangga dengan persamaan (2-4) sebagai berikut [PHD-10].

$$W_{(i)} = \frac{1}{d+a} \quad \dots(2-4)$$

Dimana  $d$  adalah jarak dan  $a$  merupakan nilai *regulator smoothing*, dimana dalam penelitian ini menggunakan  $a=0,5$ . Weight voting ini kemudian dijumlahkan setiap kelasnya dan kelas dengan jumlah terbesar yang akan dipilih menjadi sebuah keputusan. Dalam metode MKNN masing-masing  $k$  tetangga terdekat dihitung menggunakan persamaan (2-3). Setelah itu, nilai validitas setiap data yang telah dihitung sebelumnya dikalikan dengan hasil weight voting berdasarkan jarak. Sehingga dalam metode MKNN didapatkan persamaan (2-5) weight voting sebagai berikut.

$$W_{(i)} = Validitas(i) \times \frac{1}{d+0.5} \quad \dots(2-5)$$

Dimana:

$W_{(i)}$  = weight voting

Validitas(i) = nilai validitas

$D$  = jarak

Teknik weight voting ini berpengaruh terhadap data yang mempunyai nilai validitas lebih tinggi dan paling dekat dengan data. Selain itu, dengan mengalikan validitas dengan jarak data mengatasi kelemahan antara jarak setiap data dengan weight yang memiliki banyak masalah dalam outlier. Oleh karena itu, metode MKNN secara signifikan akan lebih kuat daripada metode KNN tradisional yang hanya didasarkan pada jarak [PHD-10].

### 2.12 Akurasi Sistem

Akurasi merupakan seberapa dekat suatu angka hasil pengukuran terhadap angka sebenarnya (true value atau reference value). Akurasi dapat diperoleh dari persentase kebenaran, yaitu perbandingan antara jumlah diagnose yang tepat

dengan jumlah data keseluruhan. Akurasi dinyatakan dalam persamaan(2-6) berikut.

$$akurasi = \frac{jumlah\ diagnosa\ benar}{jumlah\ data} \times 100\% \quad (2-6)$$

Dimana jumlah diagnose benar adaah jumlah record data uji yang diprediksi dimana kelasnya menggunakan metode kasifikasi dan hasilnya sama dengan kelas sebenarnya. Sedangkan jumlah data adalah jumlah keseluruhan record yang diprediksi kelasnya (seluruh data uji). Metode klasifikasi berusaha untuk mencari model yang memiliki tingkat akurasi yang tinggi ketika model tersebut diterapkan pada data uji [STD-12].

