

**DETEKSI TANGGAL KADALUARSA PRODUK
MENGUNAKAN PUSTAKA TEKNOLOGI OCR
(OPTICAL CHARACTER RECOGNITION) DAN
KLASIFIKASI K-NEAREST NEIGHBOUR (KNN)**



Diusulkan oleh :

Fitri	G64110006
Herdi Agusthio	G64110024
Hilda Sudawani	G64110090
Maulana Rizal Ibrahim	G64110101
Miqdad Aufa	G64124068

**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2015**

I. PENDAHULUAN

Latar Belakang

Dalam kehidupan sehari-hari orang sering membeli suatu produk makanan atau minuman hanya dengan memperhatikan produk tersebut sudah kadaluarsa atau belum. Biasanya pembeli tidak memperhatikan atau menghitung lama waktu sebelum produk tersebut kadaluarsa, padahal semakin sedikit rentang waktu sebelum kadaluarsa, semakin menurun pula kualitas dan keamanan dari produk. Hal ini sesuai dengan pengertian tanggal kadaluarsa yaitu tanggal yang dipilih oleh pabrik yang memproduksi suatu produk untuk menjamin potensi yang penuh dan keamanan dari produk sebelum tanggal kadaluarsanya.

Ukuran tanggal kadaluarsa yang kecil juga menyulitkan orang yang memiliki gangguan penglihatan dalam melihat tanggal kadaluarsa. Hal ini dapat menyebabkan pembeli salah melihat tanggal dan membeli produk yang sudah kadaluarsa. Oleh karena itu, dikembangkan sistem pendeteksi tanggal kadaluarsa untuk mendeteksi suatu produk sudah kadaluarsa atau belum dan menghitung lama waktu sebelum produk kadaluarsa. Output sistem akan ditampilkan dengan ukuran tulisan yang lebih besar sehingga mudah dibaca oleh pembeli.

Perumusan Masalah

Perumusan masalah dalam sistem ini mencakup beberapa hal sebagai berikut:

1. Bagaimana menerapkan metode PCA dan KNN dalam sistem ini?
2. Bagaimana tingkat akurasi dari metode yang digunakan?

Tujuan

Tujuan sistem ini adalah mengembangkan sistem pendeteksi tanggal kadaluarsa produk dengan menerapkan metode KNN sebagai metode klasifikasinya dan metode PCA untuk ekstraksi ciri.

Manfaat

Manfaat penelitian ini adalah membantu pembeli dalam menghitung rentang waktu sebelum produk kadaluarsa dan membantu pembeli yang memiliki gangguan penglihatan dalam melihat tanggal kadaluarsa. Dengan menghitung rentang waktu kadaluarsa, pembeli dapat terhindar dari membeli produk yang sudah kadaluarsa. Pembeli juga dapat membeli produk dengan rentang waktu kadaluarsa yang lebih lama sehingga kualitas produk yang dibeli pun lebih baik.

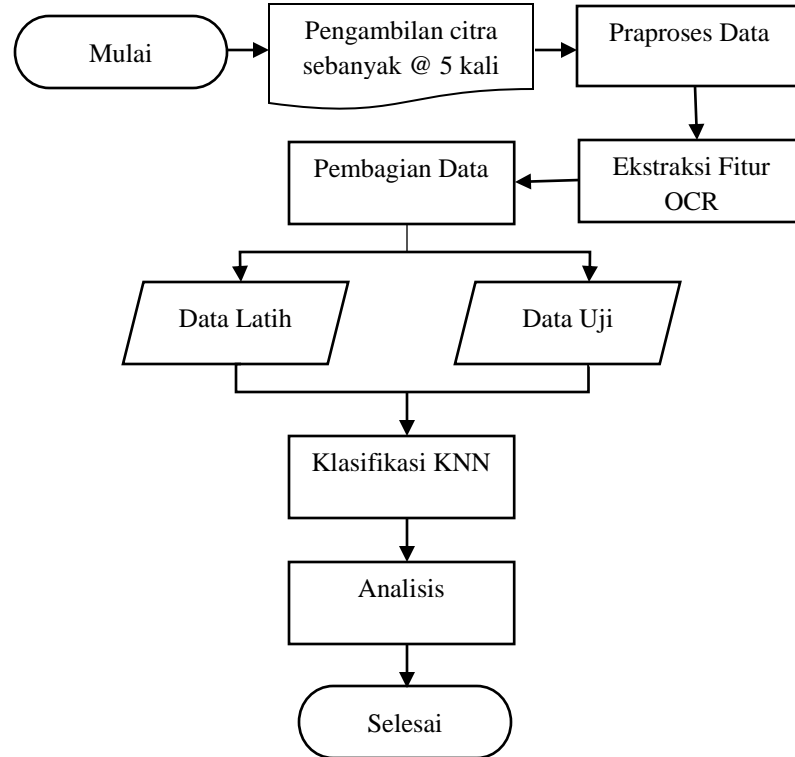
Ruang Lingkup

Ruang lingkup dari sistem ini adalah:

1. Data tanggal kadaluarsa diambil dari satu jenis produk makanan kaleng yaitu kornet Pronas.
2. Data yang diolah berasal dari citra dengan format .jpeg.
3. Ukuran citra yang diolah harus berukuran 1600 x 900 piksel.
4. Karakter yang dikenali adalah angka 0-9.
5. Pengambilan citra dilakukan pada kondisi pencahayaan tidak silau oleh lampu, jarak pengambilan 6 cm tegak lurus, pengaturan kamera *autofocus* dan latar di dalam supermarket.

II. METODE

Sistem dimulai dengan mengambil data citra, praproses data, ekstraksi fitur dengan OCR, pembagian data latih dan data uji, klasifikasi KNN, dan analisis. Ilustrasi sistem dapat dilihat pada Gambar 1.



Gambar 1 Diagram Alir Sistem

Pengambilan Data Citra

Data citra pada diambil menggunakan kamera *handphone* dengan resolusi 8 MP. Data yang digunakan sebanyak 4 citra. Data tersebut kemudian diambil potongan karakter yang mewakili angka 0-9 pada kornet Pronas. Masing-masing karakter diambil sebanyak 1 buah.

Praproses Data

Pada tahapan ini, data yang telah dikumpulkan dengan resolusi tinggi akan diubah ke resolusi lebih kecil. Data citra dari format warna RGB diubah menjadi *grayscale*. Format *grayscale* memudahkan pemrosesan data karena citra *grayscale* hanya butuh sedikit informasi yang diberikan tiap piksel. Proses konversi *grayscale* dapat dihitung menggunakan rumus:

$$\text{Pixel} = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B$$

Thresholding

Gambar hitam putih (binary image) relatif lebih mudah dianalisis dibanding gambar berwarna. Karena itu, sebelum dianalisis gambar dikonversikan terlebih dahulu menjadi binary image. Proses konversi ini disebut thresholding. Dalam proses thresholding, warna

yang ada dikelompokkan menjadi 0 (hitam) atau 1 (putih). Pengelompokannya didasarkan pada suatu konstanta ambang batas (T). Thresholding digunakan untuk mengubah gambar dengan mode grayscale atau gambar berwarna menjadi gambar hitam putih (biner) dengan tingkat kontras yang sangat tinggi. Semua pixel yang lebih terang dari threshold akan diubah menjadi putih, sebaliknya semua pixel yang lebih gelap dari threshold akan diubah menjadi hitam. Proses threshold sangat berguna untuk menentukan daerah yang paling terang dan daerah paling gelap dari sebuah gambar.

Cropping

Cropping adalah memotong satu bagian dari citra sehingga diperoleh citra yang berukuran lebih kecil. Hal ini dilakukan untuk menghindari adanya analisis di luar daerah kajian. Selain itu, dilakukan untuk memudahkan melakukan analisis citra dari daerah kajian. Pemotongan juga mengakibatkan ukuran objek menjadi lebih besar, sehingga konten yang ada (informasi berupa warna) dapat terlihat lebih jelas.

Pengenalan Karakter dengan OCR (Optical Character Recognition)

OCR atau Optical Character Recognition adalah teknologi pengenalan karakter yang ditulis atau dicetak oleh komputer. Dalam kasus ini, karakter yang akan dikenali berupa angka-angka yang merepresentasikan tanggal kadaluarsa produk kornet. Beberapa proses yang dilibatkan dalam teknologi ini adalah *photo scanning* atau pemindaian foto terhadap satu per satu karakter, analisis gambar hasil pemindaian, kemudian terjemahan gambar karakter dalam kode karakter. Kode karakter yang digunakan pada proyek PPCD ini adalah alphabet-numerik standar, bukan bilangan ASCII atau yang lainnya. Kode karakter menjadi keluaran dalam format *char* pada program. (Rouse 2005)

Dalam proses OCR, hasil pemindaian gambar dianalisa terhadap nilai gelap terangnya area dari gambar. Hal ini dilakukan untuk mengidentifikasi setiap alfabet-numerik yang ada. Setelah sebuah karakter dikenali, maka karakter tersebut dikonversikan ke dalam kode karakter. (Rouse 2005)

Pembagian Data

Data yang telah diekstraksi kemudian dibagi menjadi data latih dan data uji. Pembagian data dilakukan menggunakan metode *k-fold cross validation* dengan kombinasi $k = 1$. Proses identifikasi dilakukan dengan 1 fold. Data yang digunakan pada *1-fold cross-validation* ini adalah data latih dengan jumlah 63.

Klasifikasi KNN

Metode klasifikasi yang digunakan pada sistem ini yaitu KNN. Konsep dasar dari KNN adalah mencari jarak terdekat antara data yang akan dievaluasi dengan k tetangga terdekatnya. Nilai dari jarak antara data latih dengan data uji diurutkan dari nilai terendah. Kelas dari nilai dengan jarak terendah diperiksa, sedangkan kelas dengan nilai jarak tertinggi menjadi kelas dari data uji tersebut.

Jarak antara dua titik dalam ruang fitur dapat dicari menggunakan jarak Euclid. Hasil dari perhitungan jarak Euclid digunakan untuk menentukan kemiripan antara data latih dan data uji. Kecocokan dilihat dari nilai (jarak) yang paling minimum. Jarak Euclid dihitung dengan rumus berikut:

$$D_{(p,q)} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

$D(p,q)$ merupakan jarak antara vektor p yang merupakan sebuah titik yang telah diketahui kelasnya dan q yang merupakan titik baru yang merepresentasikan data yang akan dijadikan data uji. Jarak antara vektor dan titik dari data latih akan dihitung dan diambil k buah vektor terdekat.

Tahapan algoritma KNN adalah sebagai berikut:

1. Menentukan parameter k (jumlah tetangga terdekat).
2. Menghitung jarak antara data yang masuk dan semua sampel latih yang sudah ada menggunakan jarak Euclid.
3. Tentukan k label data yang mempunyai jarak yang minimal.
4. Klasifikasikan data baru ke dalam label data yang mayoritas.

Analisis

Setelah diperoleh output dari KNN, maka akan dihitung akurasi tiap kelasnya dengan persamaan berikut:

$$\text{akurasi} = \frac{\sum \text{data uji benar}}{\sum \text{data uji}} \times 100\%$$

Lingkungan Pengembangan

Lingkungan yang digunakan untuk sistem ini memiliki spesifikasi sebagai berikut:

Perangkat keras:

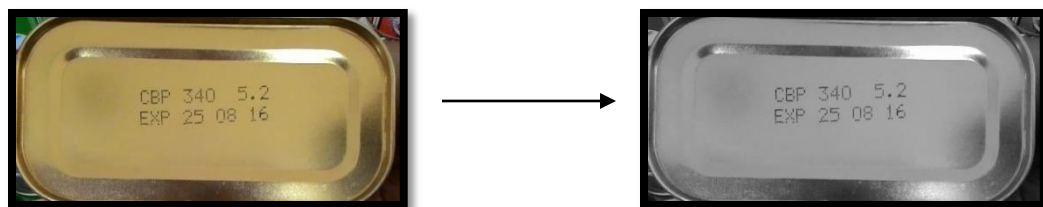
1. Processor Intel Core i5 2.40 GHz
2. RAM 4 GB
3. Harddisk 500 GB

Perangkat lunak:

1. Sistem operasi Windows 7 Ultimate 32-bit
2. Codeblocks
3. Adobe Photoshop CS3

III. HASIL DAN PEMBAHASAN

Pada praproses citra, data yang telah dikumpulkan akan dilakukan teknik *grayscale*, *thresholding*, dan *cropping*.



grayscale

thresholding

cropping

Pada tahap *grayscale*, citra asli diubah menjadi citra *grayscale* untuk memudahkan pemrosesan data dan mempercepat proses komputasi. Setelah itu, dilakukan *thresholding* untuk menentukan daerah yang paling gelap dan terang dari citra. Terakhir, kemudian dilakukan *cropping* memudahkan melakukan analisis citra dari daerah citra.

Klasifikasi dengan metode KNN dilakukan pada data latih sebanyak 63 dan data uji sebanyak 4. Pembagian data dilakukan menggunakan *k-fold cross validation* dengan *k* sebesar 1. Parameter KNN adalah *data test* (masukan dari *user*), *data training*, kelas dari *data training*, dan jumlah tetangga terdekat (*k*). Rumus jarak yang digunakan yaitu jarak Euclid. Nilai *k* pada KNN yang digunakan adalah 5.

Pada tahap pengujian, data yang akan diujikan adalah 4 citra. Dari data citra tersebut, akan dinilai akurasi untuk mengetahui hasil akhir yang didapat dari metode KNN. Rumus untuk mengukur akurasi yaitu :

$$\text{Akurasi} = \frac{\text{jumlah karakter yang dikenali dengan benar}}{\text{total karakter}} \times 100\%$$

Berikut tabel pengujian untuk citra yang sudah dilakukan *cropping* secara otomatis :

No.	Gambar kadaluarsa kornet	tanggal produk	Nilai	Pengenalan	Akurasi (%)
1.	CBP 340 3.2 EXP 25 08 16		EXP 25 08 16	T 7 j j j 1 j j I j	0 %
2.	CBP 340 4.1 EXP 16 09 16		EXP 16 09 16	j 0 1 j 0 9 3 j 1 1 0 9 1 6	83.3%
3.	CBP 198 5.3 EXP 11 09 16		EXP 11 09 16	j j 1 j L 6 j j j 1 1 Q 7 I Q	33.33%
4.	CBP 198 4.1 EXP 10 09 16		EXP 10 09 16	0 0 0 1 0 4 0 1 0 0 0 1 8	66.66%

Tabel Pengujian Tanggal Kadaluarsa Produk Kornet

Akurasi hasil pengujian untuk citra yang dilakukan cropping secara otomatis tidak terlalu baik yaitu 45.83%. Tetapi untuk citra yang dilakukan cropping secara manual, akurasi jauh lebih baik. Berikut citra yang sudah terdeteksi secara tepat tanggal kadaluarsanya dan sisa waktu sampai produk kadaluarsa.

```
Deteksi Tgl Kadaluarsa : Metode PCA dan KNN : PPCD
tanggal exp = 25
bulan exp = 8
tahun exp= 2016
Sisa waktu sampai produk kadaluarsa 568 hari
```

KESIMPULAN

Berdasarkan hasil analisis dan pengujian yang dilakukan, kesimpulan yang dapat diambil adalah :

1. Sistem secara keseluruhan dapat mendeteksi dan mengenali tanggal kadaluarsa produk kornet Pronas dengan akurasi jumlah 45.83% (4 kornet) dengan jumlah database 63 jumlah karakter untuk pengujian sebanyak 4 kornet.
2. Algoritme pendekteksian tanggal kadaluarsa produk kornet Pronas dipengaruhi oleh beberapa faktor :
 - Tingkat kecerahan
 - Kemiringan citra karakter
 - Teknik *cropping*

DAFTAR PUSTAKA

- Lahmurahma HF. 2013. Perbandingan pengenalan karakter plat nomor menggunakan *image centroid and zone & knn* [skripsi]. Bogor (ID): Institut Pertanian Bogor.
- Rouse, Margaret. 2005. *IT Acronyms at Your Fingertips*. Dari TechTarget. [Online]. (<http://searchcontentmanagement.techtarget.com/definition/OCR-optical-character-recognition>, diakses tanggal 19 Januari 2015).
- Saputro TA, Dhanny E, Ramadhan A, Situmorang AM, Lazuardi MF. 2014. Deteksi posisi plat nomor kendaraan bermotor berdasarkan area citra. Malang (ID) : Universitas Brawijaya.
- Taufiqurrohman, Hidayatno A, Zahra AA. 2013. Pengenalan plat nomor sepeda motor dengan menggunakan metode jarak euclidean. semarang (ID): Universitas Diponegoro.
- Yulida S, Kusumawardhan A, Setijono H. 2013. Perancangan sistem pengenalan plat nomor kendaraan menggunakan metode principal component analysis. TEKNIK POMITS. 2(1): 178.