

Image Searching Data Gambar Berwarna Algoritma K Means Clustering Pada Data Penjualan

Ariadi Retno Hayati

Politeknik Negeri Malang, Teknologi Informasi, Malang, Indonesia
E-mail: faniri4education@gmail.com

Abstrak

Pencarian data dengan data gambar berwarna yang diterapkan pada penelitian ini adalah data penjualan dengan dataset gambar berwarna. Terdapat beberapa produk yang akan dideskripsikan yaitu produk makanan, produk bahan pokok, produk rumah tangga yang akan dikelompokkan berdasarkan kategori kelompok dengan kategori produk yang sama. Metode yang diterapkan adalah *K Means Clustering* pada penelitian ini dengan data pembelajaran dan data uji coba yang berbeda. Sistem pencarian gambar dengan memilih file gambar dari data uji coba dan sistem akan mencari kelompok dari data pembelajaran dan ditampilkan gambar hasil pencarian untuk mempermudah pengguna aplikasi mencari data penjualan jika tidak memiliki atau kesulitan mencari informasi nama produk atau kode produk. Proses pengolahan data warna baik data gambar berwarna pada data pembelajaran dan data uji coba harus melalui proses normalisasi data warna dimana proses ini meliputi pemilihan objek, pemotongan objek, dan menyamakan ukuran gambar objek. Hasil pembelajaran pengelompokan *K Means Clustering* mendapatkan hasil yang baik, pada penelitian ini mendapatkan prosentase 93,33% pada data pembelajaran dan 60% pada data uji coba.

Kata kunci: Image Searching, K Means Clustering, Scilab, Data Gambar Warna Produk Penjualan

Abstract

Search engine with image data in this research use color image for sale data, the purpose of this research is to cluster data as the input data clustering. The result from clustering with K Means Clustering method is data with similar characteristic with the input color image to help user application search data without know information name and code product There is a view product which separated as different classes. Each class has ten data and use for learning data five data, for testing data five data. Each class has different data and build from data sales like soap product, food product, and homemade product. In this research build application for analyze result where input data choose from name file and then process in K Means clustering with learning data and get the cluster value for each class and then view the learning data which clustered for data input in learning system. Before data in processing in K-Means Clustering data should normalized process as resize size image in same size and then change rgb value image between zero and one value. Percentage result is 93,33% for learning and testing is 60 %.

Keywords: Image Searching, K Means Clustering, Scilab, Colour Image for Sale Product

1. PENDAHULUAN

Metode *K Means Clustering* diterapkan untuk mengelompokkan data secara *unsupervised*, pada penelitian ini metode *K Means Clustering* diterapkan untuk *Image Searching* pada data penjualan produk. Penerapan metode *K Means Clustering* pada gambar pada penelitian sebelumnya sudah diterapkan, berikut beberapa penelitian dengan data gambar yang menerapkan metode pengelompokan data untuk *Image Searching* dan *Image Retrieval*. Pada penelitian yang berjudul *Content Based Image Retrieval in Java Using K-Means Clustering Algorithm* [7], dalam penelitian ini menerapkan data gambar sebagai *input query* yang berikutnya akan dibandingkan dengan data di basis data dan dikelompokkan dimana menggunakan bahasa pemrograman Java dengan data set dengan gambar berwarna. Menurut [13], pada penelitian dengan judul *The Application of Similar Image Retrieval in Electronic Commerce* yang menerapkan dengan dataset data elektronik dengan pengukuran kesamaan pencocokan nilai data. Menurut [9], pada penelitian dengan judul *Survey Paper on Content Based Image Retrieval*, bahwa *content* pada pembahasan *Image Retrieval* meliputi bagaimana mencari gambar dengan benar, bagaimana manajemen gambar pada sistem, dan bagaimana

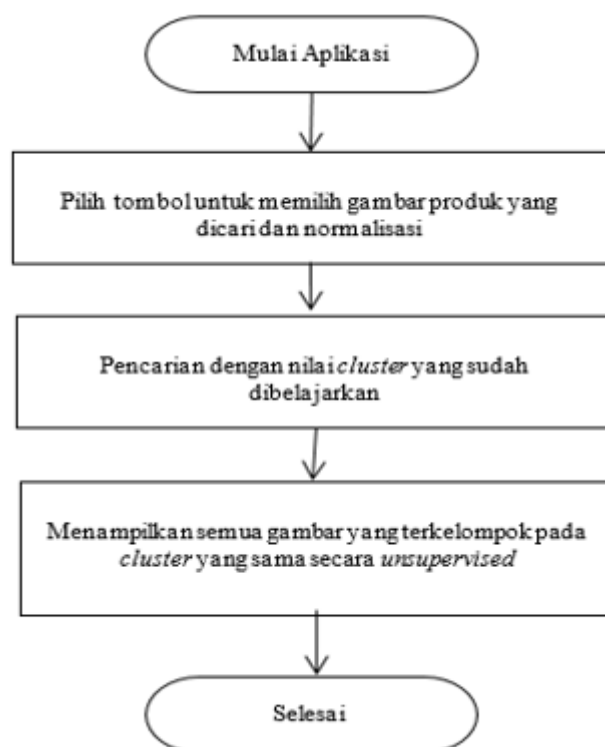
proses mencari gambar pada sistem. Menurut [5], pada penelitian dengan judul *Content Based Image Retrieval Using Clustering*, dalam penelitian ini menerapkan *Image Retrieval* dengan metode pengelompokan *clustering* dengan kernel Gaussian.

Dataset yang digunakan pada penelitian adalah dataset yang dibangun sendiri yaitu data gambar berwarna untuk produk-produk penjualan yang dapat diimplementasikan di sistem, dengan bahasa pemrograman Scilab. Metode yang digunakan pada penelitian ini yaitu menggunakan *K Means Clustering* dimana produk akan terkelompok secara *unsupervised* dengan inisialisasi *cluster*. Terdapat data khusus pembelajaran dan data untuk uji data set, dimana dengan karakter data warna diubah menjadi dua dimensi dan proses pembelajaran *K Means Clustering* mendapatkan hasil prosentase 93,33% pada data pembelajaran yang diperoleh dengan cara menghitung jarak setiap data set pembelajaran dengan hasil akhir nilai *center* dari metode *K Means Clustering* dan dianalisa apakah sesuai dengan kelompok *cluster* data yang benar sebagai sarana pembanding untuk menelaah keberhasilan aplikasi data set pembelajaran. Sedangkan pada data uji coba yaitu data yang belum diproses pada pembelajaran *K Means Clustering* mendapatkan pengenalan gambar 60% yang diperoleh dengan cara menghitung jarak setiap data uji coba dengan nilai *center* dari hasil proses pembelajaran dan dibandingkan dengan kelompok *cluster* yang benar dari data set uji coba dimana belum dibelajarkan pada metode *K Means Clustering*. Sesuai hasil prosentase dengan *unsupervised* data pengelompokan pada metode *K Means Clustering* dengan *euclidean distance* diterapkan pada produk penjualan dapat diterapkan sesuai prosentase 93,33% dengan bahasa pemrograman Scilab pada aplikasi ini.

2. METODOLOGI PENELITIAN

2.1. Konsep *Image Searching*

Image Searching memiliki konsep yang sama dengan *Text Searching* dimana perbedaan pada *query* pada *Image Searching* *query* menggunakan data gambar (*colour image*). Pada penelitian ini menggunakan dataset yang dibuat sendiri dengan data-data produk penjualan dengan pembelajaran secara *unsupervised* sebagaimana flowchart pada gambar 1.



Gambar 1. Konsep *Image Searching*

Proses pencarian menggunakan gambar mempermudah user jika tidak mengetahui nama atau merk suatu produk. Pendataan suatu produk pada umumnya menggunakan data yang diinputkan pada kolom teks pencarian data yaitu:

- a) id produk
- b) nama produk

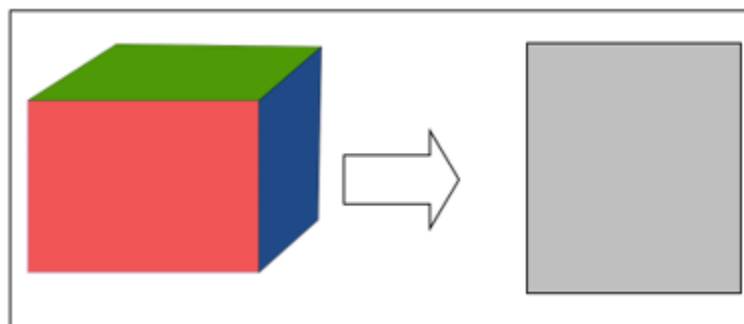
Jika pengguna tidak mengetahui id produk ataupun nama produk dapat menerapkan gambar sebagai *query* pencarian data, pada aplikasi yang diterapkan pada penelitian ini menggunakan input file dari gambar yang tersedia di folder. Menurut [4][8], hasil *query* akan ditampilkan gambar-gambar yang sesuai dengan karakteristik yang mirip dengan input *query* gambar.

2.2. K Means Clustering

Pada penelitian ini menerapkan metode *K Means Clustering* menurut [2][3], pada data gambar warna yaitu data penjualan produk produk. Rumus *K Means Clustering* sebagaimana berikut:

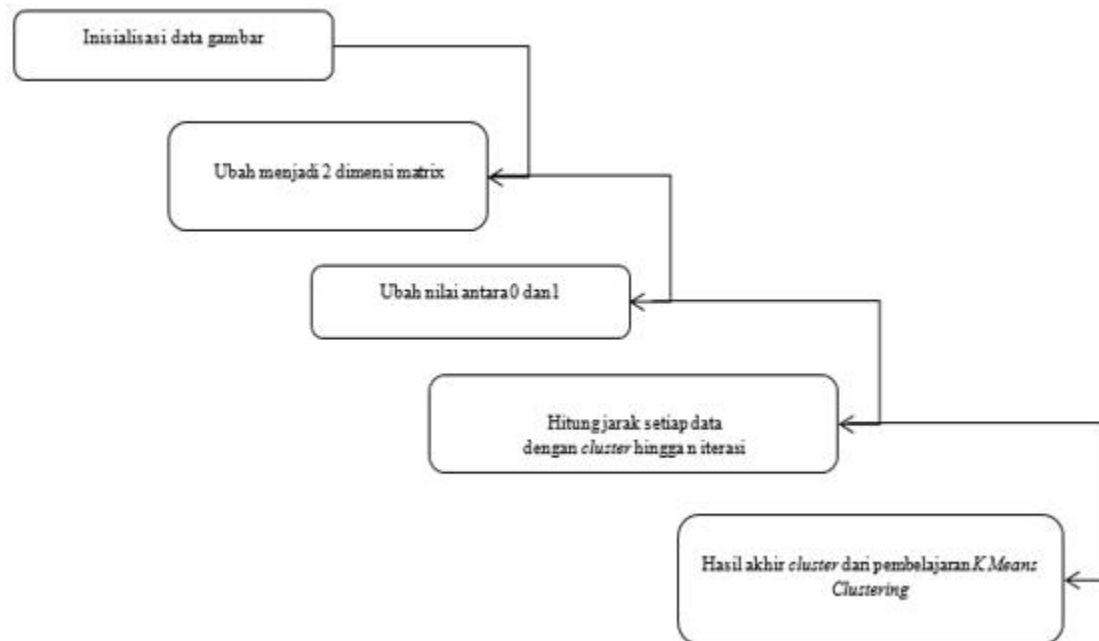
<i>Input</i>	: N Data pembelajaran
<i>Output</i>	: k <i>cluster</i> hasil pengelompokan data dengan nilai jarak terdekat
<i>Process:</i> 1. Inisialisasi nilai awal sejumlah kelompok produk misalkan ($m_1, m_2, \dots, m_k$) sesuai data pembelajaran (nilai sudah dinormalisasi), inisialisasi iterasi=1. 2. Hitung jarak setiap objek gambar produk yang sudah dinormalisasi menjadi 2 dimensi dengan <i>cluster</i> awal pada nomor 1 dengan metode Euclidian. 3. Hitung rata-rata berdasarkan jarak terdekat dengan rata-rata pada setiap dimensi data yang terkelompokkan pada m <i>cluster</i>. 4. Apakah sudah konvergen atau sesuai dengan iterasi N jika belum maka proses no 2 dan 3 diproses hingga sesuai iterasi 5. Jika sudah hingga iterasi N maka algoritma selesai.	

Menurut [10] [11] [12], proses *K Means Clustering* pada d dimensi memiliki proses yang sama dengan *high dimensional data*, tetapi pada *high dimensional data* membutuhkan jumlah perulangan sejumlah dimensi data. Pada penelitian ini data gambar terdiri dari 3 dimensi yaitu *red* dimensi, *green* dimensi, *blue* dimensi diubah menjadi dua dimensi yaitu berwarna abu-abu.



Gambar 2. Mengubah 3 Dimensi Menjadi 2 Dimensi

Diagram Pembelajaran *K Means Clustering*

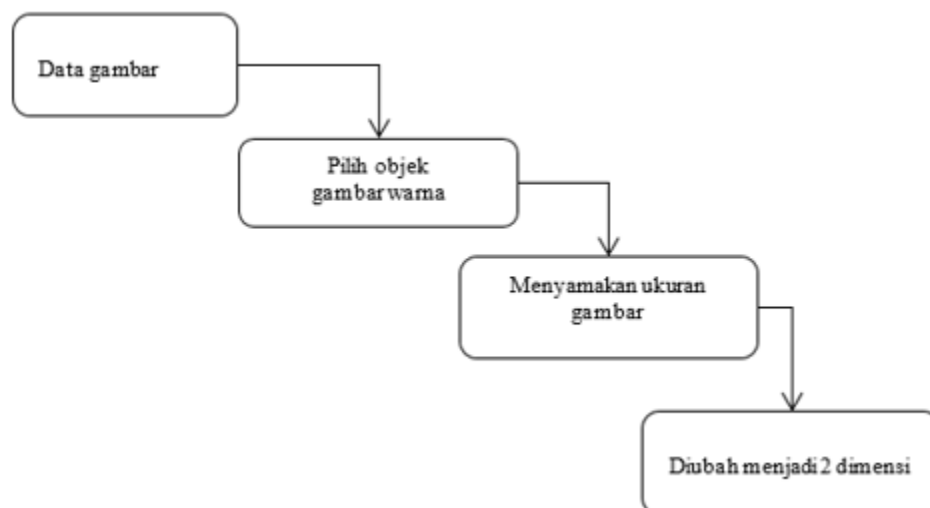


Gambar 3. Diagram Pembelajaran K Means Clustering

Pada gambar 2 merupakan ilustrasi pengubahan data warna yang memiliki data matriks berwarna merah (*red*), hijau (*green*), biru (*blue*) diubah menjadi data abu abu (*gray*) yang memiliki desain 2 dimensi matriks. Pada gambar 3, merupakan bentuk diagram pembelajaran *K Means Clustering* yaitu menghitung jarak dan pengelompokan hingga n iterasi pada akhirnya menyimpan nilai *mean* dari hasil pembelajaran *K Means Clustering*.

2.3. Data Set

Diagram Normalisasi data pada gambar 4, dimana terdapat proses keseragaman data dan pengubahan dimensi pembelajaran sesuai [6], untuk normalisasi data.



Gambar 4. Diagram Proses Normalisasi Data

Pada proses normalisasi sesuai diagram, merupakan proses menyamakan ukuran gambar objek produk yang terpilih, mengubah menjadi dua dimensi dengan warna *gray* dan mengubah angka 0-255 menjadi antara 0 dan 1.



Gambar 5. Beberapa Data Set Produk Penjualan

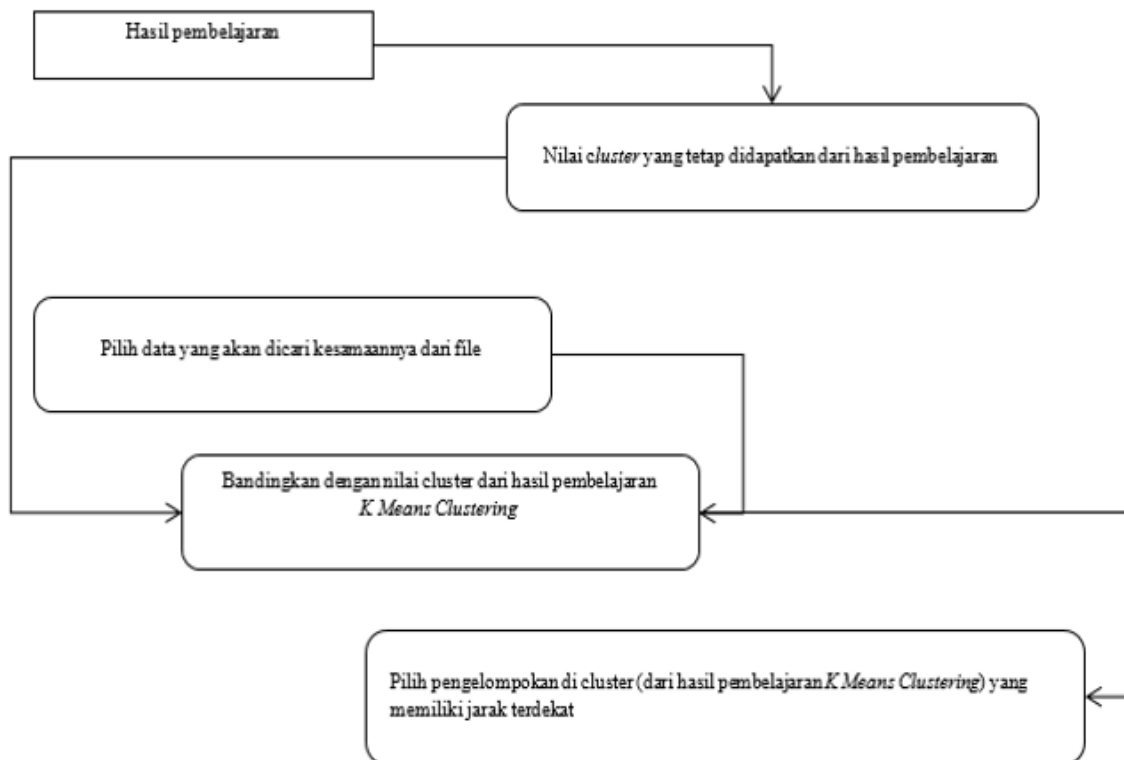
Pada gambar 5 merupakan beberapa data produk yang digunakan sebagai dataset pembelajaran dan uji coba produk penjualan memiliki tekstur yang berbeda pada setiap data, kemiripan diutamakan dari karakteristik warna.

2.4. Tahap Penelitian

Data pembelajaran dan data uji atau testing, untuk penambahan data pembelajaran dapat ditambahkan di data set data pembelajaran dan akan diproses kembali dengan metode *K Means Clustering* dan menghasilkan nilai dengan feature dimensi yang berbeda. Penjelasan dari gambar 6 adalah setelah mendapatkan nilai *center* dari data pembelajaran selanjutnya data-data warna yang akan diujicobakan pada sistem diproses pada setiap data dan dikelompokkan pada *cluster* dengan jarak yang terdekat tanpa mengetahui data pengetahuan yaitu data kelompok yang benar dari data tersebut. Untuk menghitung prosentase pada data tidak diketahui pada data pembelajaran dan uji coba dengan cara menghitung hasil pengelompokan data misalkan terkelompokkan pada kelompok 1 pada hasil uji coba, akan dibandingkan dengan data kelompok yang sebenarnya, jika nilainya sama yaitu terkelompokkan pada kelompok 1, maka jumlah data keberhasilan bertambah sebagai prosentase keberhasilan pengelompokan data *K Means Clustering*.

Diagram tahapan pada penelitian ini menurut [14], sebagaimana berikut:

- Tahap Pengambilan Data dengan pembuatan dataset penjualan.
- Desain *user interface* pada proses pencarian data dan penginputan data.
- Pengembangan dengan pembuatan program pada penelitian ini dengan bahasa pemrograman scilab.
- Penerapan *user interface* dalam bentuk aplikasi *image searching* data penjualan.
- Uji coba keberhasilan dengan prosentase hasil pembelajaran dan uji coba.

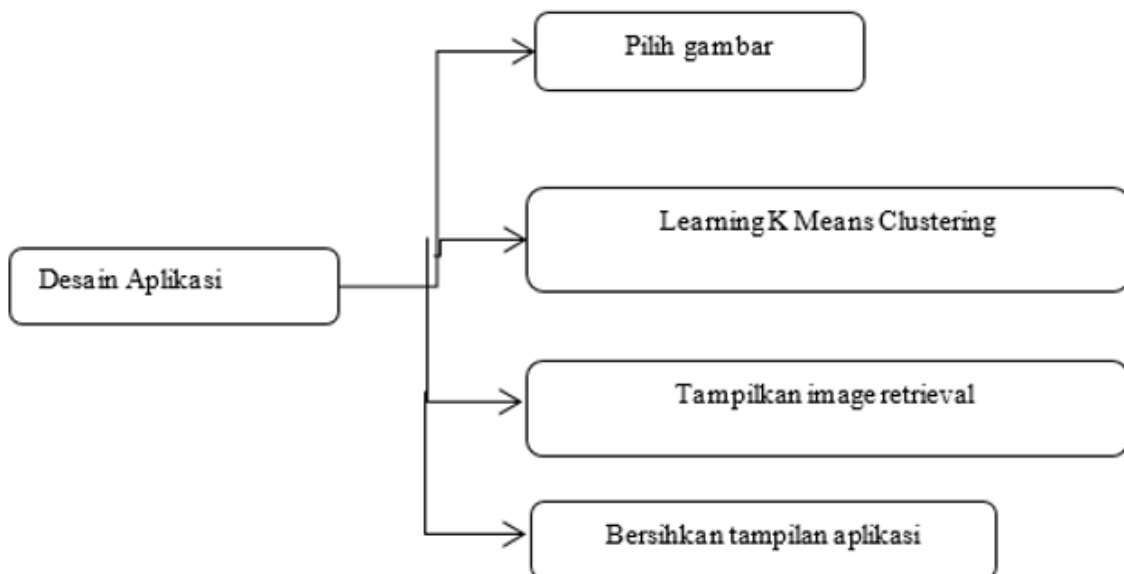


Gambar 6. Diagram Tahapan Penelitian Image Searching Metode K Means Clustering

3. HASIL DAN PEMBAHASAN

3.1. Desain Aplikasi Sistem

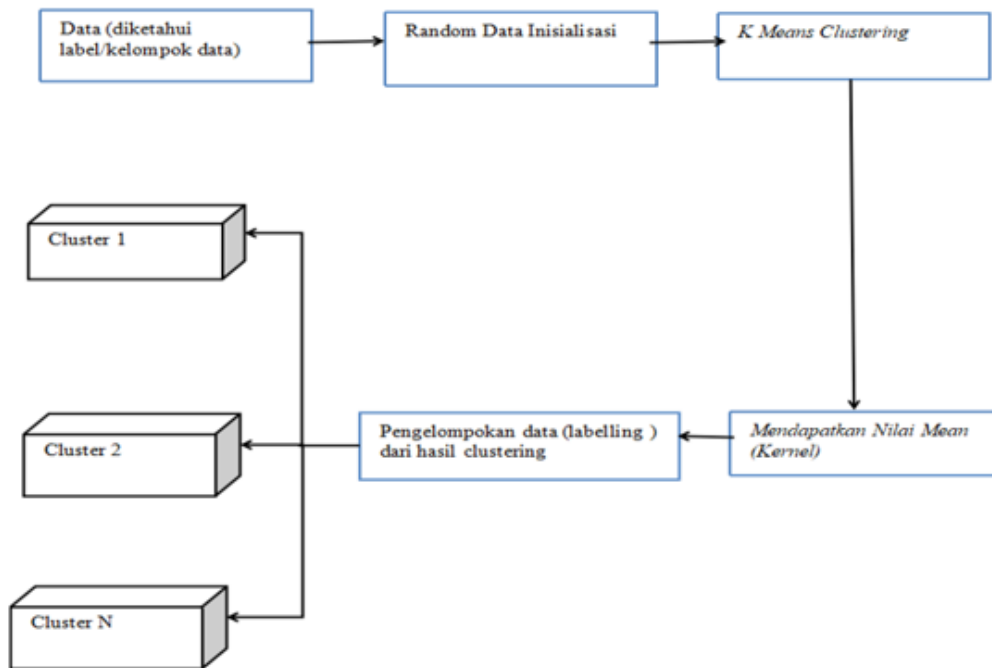
Desain aplikasi pada penelitian ini sebagaimana gambar 7, dimana terdapat beberapa pilihan yaitu memilih gambar yang akan diuji, proses pembelajaran *K Means Clustering*, Menampilkan gambar sesuai *cluster* dan membersihkan layar dengan pemrograman scilab menurut [1], sebagai pengolahan pada data gambar.



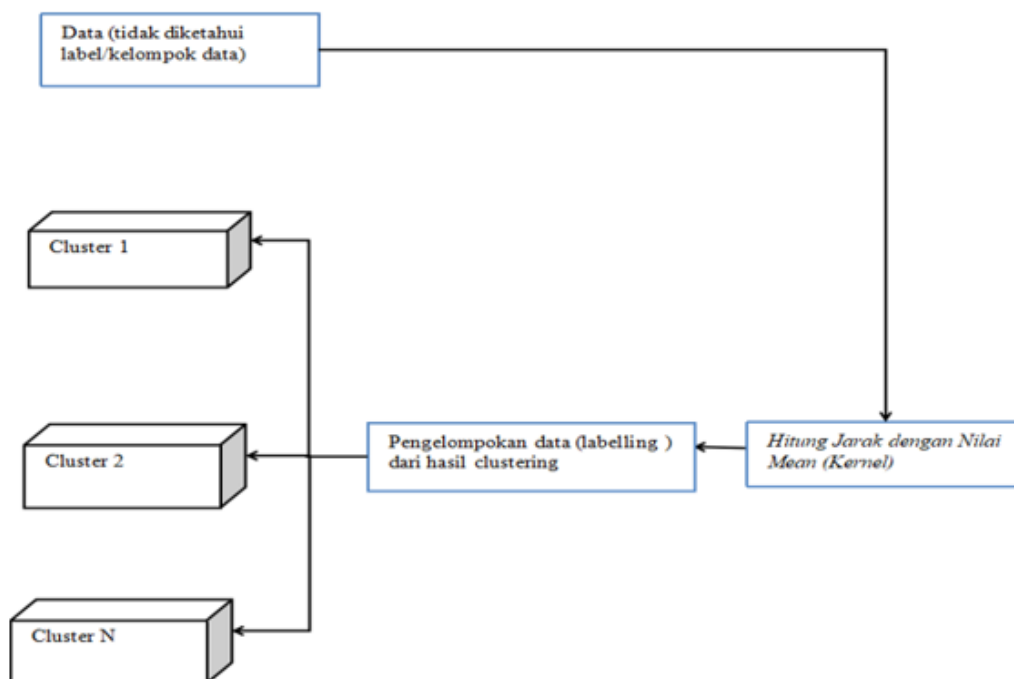
Gambar 7. Diagram Desain Aplikasi

3.2. Desain Susunan Fungsi Pembelajaran dan Uji Coba

Diagram alur sistem pada aplikasi yang diterapkan pada proses pembelajaran dan uji coba pada penelitian ini sebagai berikut:



Gambar 8. Block Diagram Proses Pembelajaran



Gambar 9. Block Diagram Proses Uji Coba

Pada proses pembelajaran sebagaimana gambar 8 dan uji coba sebagaimana pada gambar 9 terdapat beberapa perbedaan, yaitu:

- Pada data pembelajaran data diketahui *labelling* atau kelompok data untuk mengetahui prosentase keberhasilan data pembelajaran.
- Terdapat proses random data untuk inisialisasi awal
- Terdapat perbaikan nilai center/mean dengan *K Means Clustering*
- Menyimpan nilai mean dari hasil pembelajaran
- Pengelompokan data pembelajaran (*Cluster 1 hingga N*)

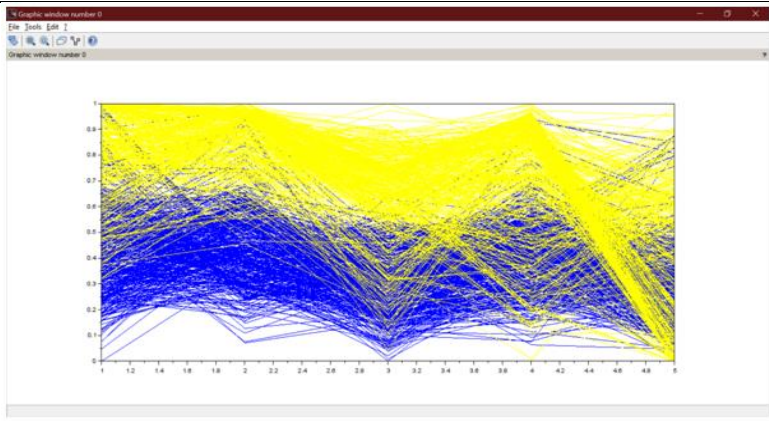
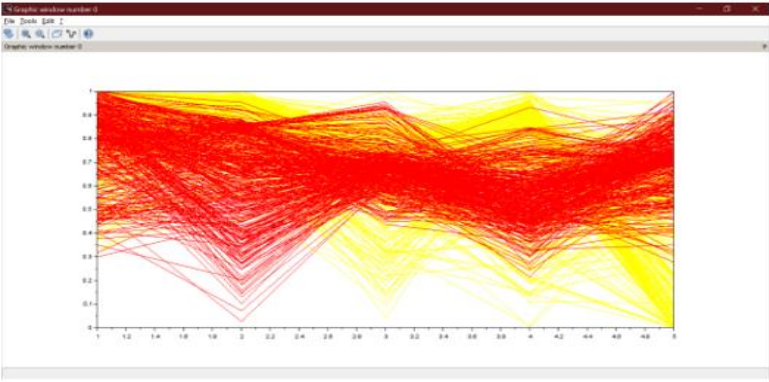
Pada proses uji coba, terdapat perbedaan jika dibandingkan dengan tahapan pembelajaran yaitu:

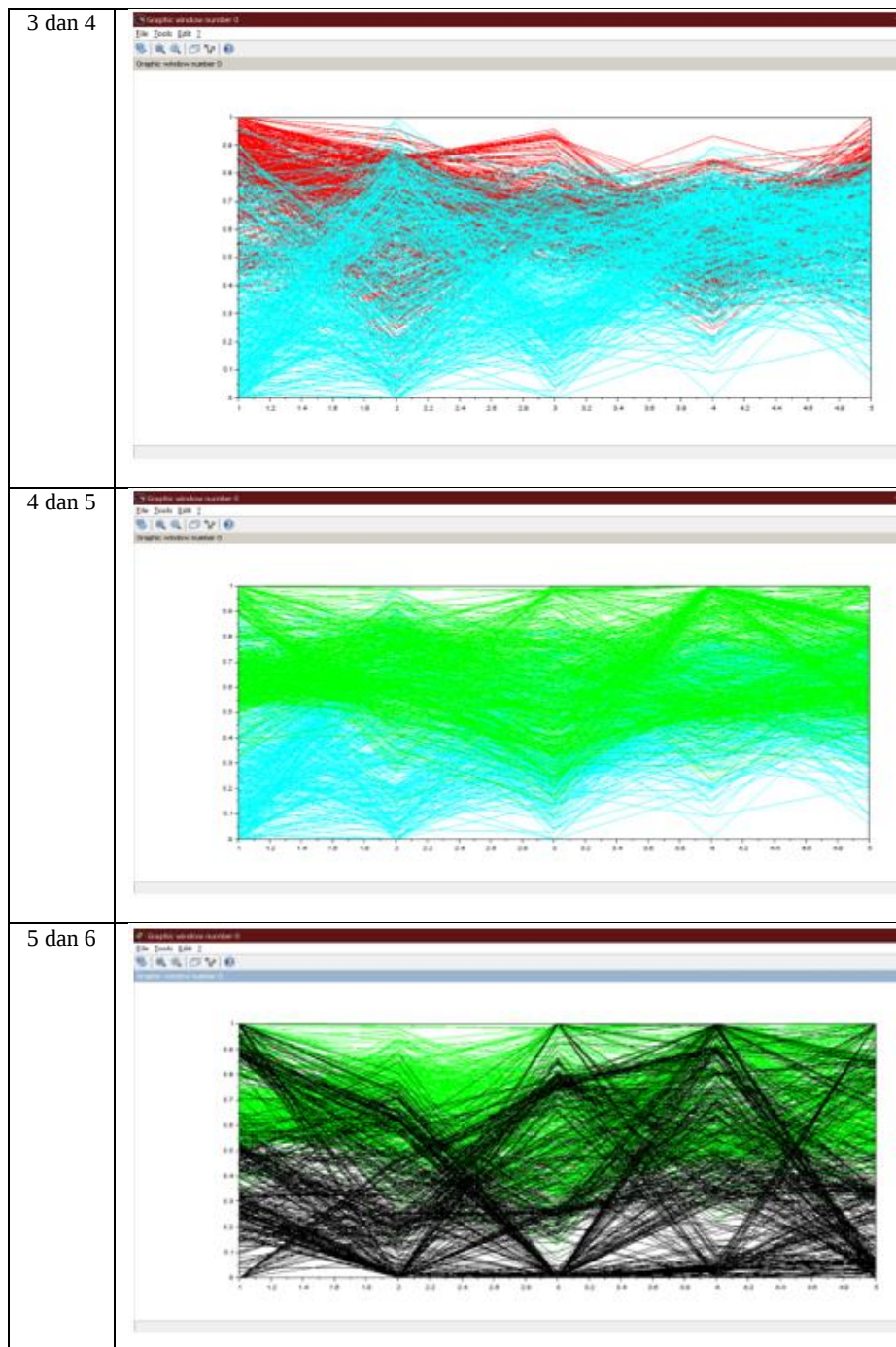
- Pada uji coba data tidak diketahui kelompoknya (*unlabelled*)
- Tidak terdapat perbaikan iterasi nilai mean (tanpa *cross validation*)
- Nilai mean dari proses pembelajaran
- Pengelompokan dengan menghitung jarak terdekat dari cluster 1 hingga N

Penghitungan prosentase keberhasilan dengan cara membandingkan data uji coba dengan data pengetahuan kelompok agar mendapatkan pengetahuan keberhasilan data jika sesuai.

3.3. Analisis Karakteristik Data Yang Digunakan Penelitian Aplikasi

Tabel 1. Karakteristik Data

Cluster	Karakteristik data
1 dan 2	
2 dan 3	



Jumlah data adalah 30 data pembelajaran dengan normalisasi data menjadi 400 dimensi yang ditampilkan sebagaimana tabel1. Baris menunjukkan data dan kolom menunjukkan dimensi, karakteristik data diubah dari *red green blue (rgb)* menjadi *grayscale*, data yang digunakan sudah ternormalisasi dengan ukuran data yang sama. Nilai dimensi pada beberapa nilai dimensi gambar pada data terdapat nilai yang memiliki karakter yang berbeda dan hal ini sangat mempengaruhi pada proses penghitungan jarak pada metode K Means Clustering, walaupun pada beberapa nilai dimensi terdapat beberapa nilai dimensi yang mirip.

70	66	7
46	8	48
5	48	67

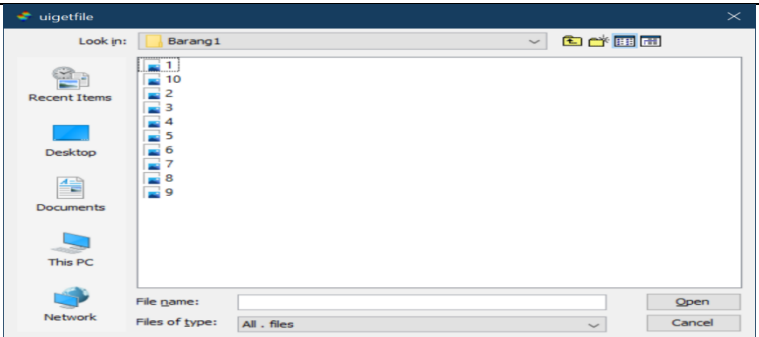
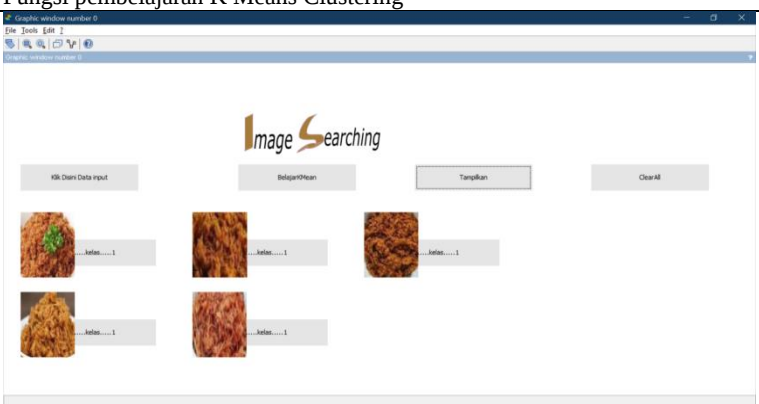
Gambar 10. Ilustrasi Desain Data Pada Pixel Dimensi Gambar

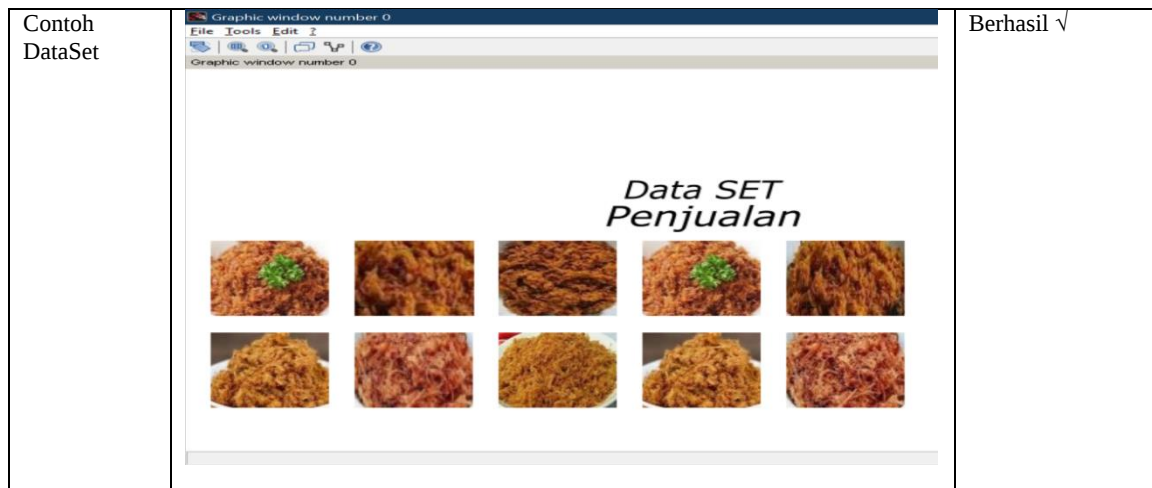
Misalkan pada contoh gambar 10, data *cluster* 1, memiliki kecenderungan nilai antara 60 hingga 90 pada dimensi 1 dan nilai antara 40 hingga 50 pada dimensi 2, dan antara 0 hingga 30 pada dimensi 3. Pada data *cluster* 2 nilai dimensi antara 0 hingga 30 terletak pada dimensi kedua, dan pada *cluster* 3 pada dimensi pertama.

3.4. Pengujian Image Searching Dengan K Means Clustering

Pada tabel 2, merupakan contoh penerapan aplikasi dimulai dari memilih data, hingga proses menampilkan data.

Tabel 2. Aplikasi

Hasil Uji Coba	Tampilan	Hasil
Pilih Gambar		Berhasil ✓
Pembelajaran	Fungsi pembelajaran K Means Clustering	Berhasil ✓
Menampilkan		Berhasil ✓
Bersihkan Aplikasi	Fungsi membersihkan gambar	Berhasil ✓



Metode *K Means Clustering* pada penelitian ini diterapkan bertujuan untuk mengelompokkan data gambar warna (*colour image*) dengan data set penjualan barang dari aplikasi dengan tujuan mempermudah user jika kesulitan

1. Tidak mengetahui kode barang atau **kesulitan mengetahui kode produk**
2. Tidak mengetahui atau **kesulitan mengetahui nama produk**

Image searching sebagai solusi permasalahan ini dengan cara:

1. Inputkan gambar warna (*image colour*)
2. Pilih tombol mencari
3. Menghasilkan data gambar yang berkaitan dengan input produk sebagaimana contoh pada tabel 2.

3.5. Pengujian Hasil Pembelajaran Dengan *K Means Clustering*

K Means Clustering pada data pembelajaran menghasilkan nilai center, dan menghasilkan kelompok sebagaimana tabel 3 sebagai hasil pengelompokan *unsupervised K Means Clustering* sebagaimana tabel 3:

Tabel 3. Hasil Unsupervised K Mean Clustering

Cluster ke	Data <i>Unsupervised</i> Pembelajaran
1	1,2,3,4,5
2	6,7,8,9,10,11
3	12,13,15
4	14,16,17,18,19,20
5	21,22,23,24,25
6	26,27,28,29,30

Hasil pembelajaran *K Means Clustering* dari penelitian ini dapat mengelompokkan data-data dari data set pembelajaran dengan data yang beragam (sudah dinormalisasi) dan menghasilkan pengelompokan data sebagaimana tabel 3. Pada kelompok 1 menghasilkan pengelompokan pada *cluster* 1 gambar 1,2,3,4,5 dimana pada dataset diketahui memiliki kelompok yang sama. *Cluster* 2, juga menghasilkan pengelompokan pada kelompok yang sama dengan pembelajaran *unsupervised*, pada *cluster* 3 data gambar dengan *cluster* yang sama hanya mengelompokkan 3 gambar dan data yang seharusnya terkelompokkan pada kelompok 3 ini

terkelompokkan pada *cluster* 4 sebagai data yang salah. Pada *cluster* 5 data gambar sudah terkelompokkan pada *cluster* yang sama, dan *cluster* 6 juga terkelompokkan pada *cluster* yang sama sebagai pembelajaran *Unsupervised*.

3.6. Analisis Kesalahan Aplikasi



Gambar 11. Grafik Kesalahan Setiap Iterasi

Hasil grafik kesalahan dari *K Means Clustering* digambarkan pada gambar 11 dan mengalami penurunan kesalahan hal ini menunjukkan adanya perbaikan nilai center pada pembelajaran *K Means Clustering* setiap iterasi. Pengelompokkan data berdasarkan penghitungan jarak dengan rumus *euclidean* untuk data berwarna pada data penjualan yang ternormalisasi dapat mengelompokkan secara *unsupervised* dengan nilai yang menghasilkan jarak terdekat.

4. KESIMPULAN

Pada penelitian ini menerapkan metode *K Means Clustering* pada data warna untuk aplikasi *Image Searching* pada *Image Retrieval* dengan pemrograman scilab, keberhasilan prosentase pada data pembelajaran dan data uji sangat dipengaruhi oleh karakteristik *feature* yang dipengaruhi oleh proses *cropping* untuk pembuatan dataset dan pengubahan dari nilai *grayscale* menjadi nilai desimal antara 0 dan 1. Penghitungan jarak dengan menggunakan metode *euclidean distance* memiliki pengaruh penghitungan jarak yang sudah tepat untuk menentukan jarak yang terdekat. Prosentase sangat dipengaruhi oleh proses normalisasi yaitu *feature data* dan pemilihan data inisial yaitu awal pada proses *K Means Clustering*.

Pada penelitian berikutnya dapat mengembangkan aplikasi penelitian ini dengan mengembangkan pada

1. Pengembangan pada proses normalisasi data dengan keseragaman proses *image processing*.
2. Pengembangan pembelajaran dengan penambahan metode agar lebih efisien.
3. Pengembangan aplikasi yang dikerjakan dengan bahasa pemrograman yang lain.

DAFTAR PUSTAKA

- [1] Singh, Laxman, and Joshi, Mauli, Face Detection System Using Scilab Image Processing Tool, *IJETCR*, vol. 3, no. 3, pp.84-90, 2015.
- [2] Cassim Alibuhtoa, Mohamed, and Idayu Mahat, Nor., Distance Based K-Means Clustering Algorithm for Determining Number of Clusters for High Dimensional Data, *Decision Science*, vol. 9, no.1, pp.51-58, 2020.
- [3] Chandra, Bhushit., Mandhaniya, Khushboo., Proposed Algorithm for Content Based Image Recognition Using Enhanced K-Means Clustering Algorithm, *IJESRT*, vol. 6, no. 10, pp. 184-190, 2017.
- [4] Sharma, Anamika., Sharma, Sarita., A Framework for Picture Extraction on Search Engine Improved and Meaningful Result, *IJCSI*, vol. 8, no. 5, pp. 1-5, 2011.
- [5] Chavan, Urvashi., Shahane, N.M., Content Based Image Retrieval Using Clustering, *IJAIE*, vol. 3, no. 10, pp. 181-184, 2014
- [6] Ramteke, Komal., Bhad, Ashwini Vinayak., Feature Extraction Approach for Content Based Image Retrieval, *IJARCC*, vol. 4, no. 11, pp. 63-68, 2015.
- [7] Kamshetty, Shradha., Kasar, Shivani., Kadu, Mrunalini., Nanajkar, J.A., *Content Based Image Retrieval in Java Using K-Means Clustering Algorithm*, *IJCRT*, vol. 6, no. 2, pp. 1492-1497, 2018.
- [8] Mundada, Ritu A., Waghmare, Akash D., An Approach for Content Based Image Retrieval Using Similarity Measures, *IJCST*, vol. 4, no. 3, pp. 124-128, 2016.
- [9] Moriwal, Rahul, Survey Paper on Content Based Image Retrieval, *IJSRD*, vol. 7, no. 7, pp. 190-193, 2019.
- [10] Wu, Chunqiong., Yan, Bingwen., Yu, Rongrui., Yu, Baoqin., Zhou, Xiukao., Yu, Yanliang., Chen, Na., K-Means Clustering Algorithm and Its Simulation Based on Distributed Computing Platform, *HINDAWI*, vol. 1, pp.1-10, 2021.
- [11] Ismia, Dewi Pramudi., Panchoob, Shireen., Murintoc, K-Means Clustering Based Filter Feature Selection on High Dimensional Data, *International Journal of Advances in Intelligent Informatics*, vol. 2, no. 1, pp. 38-45, 2016.
- [12] Umesh., Suresha., Advanced Web Image Retrieval Using Clustering Algorithms, *IJMA*, vol. 3, no. 4, pp. 159-169, 2011.
- [13] Hu, YuPing., Yin, Hua., Han, Dezhi., Yu, Fei., The Application of Similar Image Retrieval in Electronic Commerce, *HINDAWI*, pp.1-7, 2014.
- [14] Irianto, Eko., Sensune, Dana Indra., Sistem Monitoring Berbasis Project Management: PT BFI Finance Indonesia, Tbk, *Prosiding Seminar Nasional Multidisiplin Ilmu Universitas Budi Luhur*, ISSN: 2087 – 0930.