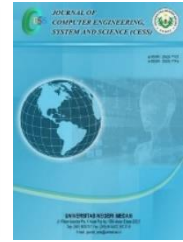


Contents list available at [www.jurnal.unimed.ac.id](http://www.jurnal.unimed.ac.id)

**CESS**  
**(Journal of Computing Engineering, System and Science)**

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



**Membandingkan Analisa Kesalahan Metode *K-Means Clustering* dan *Canopy K-Means Clustering* Dengan Data Gambar Terfilter**

***Error Analysis Comparison for Images Data with K-Means Clustering and Canopy K-Means Clustering Use Filtered Image Data***

Ariadi Retno<sup>1\*</sup>, Wilda Imama<sup>2</sup>, Puspa Kirana<sup>3</sup>, Vit Zuraida<sup>4</sup>

<sup>1,2,3,4</sup> Politeknik Negeri Malang  
Soekarno Hatta 9

email: [faniri4education@gmail.com](mailto:faniri4education@gmail.com)

**ABSTRAK**

Pada penelitian ini menganalisa kesalahan yang diperoleh pada data pembelajaran pada data gambar dengan membandingkan metode K-Means Clustering dan metode Canopy K-Means Clustering. Data yang digunakan adalah data gambar yang diujikan pada aplikasi yang dibangun dan menelaah nilai kesalahan pada setiap iterasi. Analisa kesalahan dengan memahami karakteristik formula pada metode K-Means Clustering dan Canopy K-Means Clustering dan menganalisa angka kesalahan berdasarkan formula kedua metode dengan demikian maka karakteristik perolehan error pada metode Canopy K-Means Clustering diperoleh berdasarkan karakteristik formula dari metode tersebut. Dari hasil beberapa uji coba dengan dataset yang data berbeda maka diperoleh rata-rata bahwa metode Canopy K-Means Clustering memiliki nilai kesalahan lebih sedikit sejumlah 0,0264% dibandingkan metode K-Means Clustering dengan *Euclidean distance* dan rata-rata keberhasilan 85% sesuai kelompok.

**Kata Kunci:** *Canopy K-Means Clustering, K-Means Clustering, Flowchart, Program PHP, Data Set Gambar, Filter Gambar.*

**ABSTRACT**

In this research analysis the error value from data learning based on image data compare two methods that K-Means Clustering and Canopy K-Means Clustering. Data in this research uses images data that are tested to sistem from dataset with different data each tested and compare the error result each iteration. Error result value influenced by statistic math formula from each clustered method analyzed, so can get characteristic error value from Canopy K-Means Clustering. From few test, the average from Canopy K-Means Clustering compare with K Means Clustering is 0,0264% less than error average from K-Means Clustering use *Euclidean distance* and average percentage is 85% true as the cluster.

\*Penulis Korespondensi:  
email: [faniri4education@gmail.com](mailto:faniri4education@gmail.com)

**Keywords:** Canopy K-Means Clustering, K-Means Clustering, Flowchart, PHP Programming, Image Data Sets, Image Filter.

---

## 1. PENDAHULUAN

Pada penelitian ini menelaah formula yang terdapat pada metode Canopy K Means Clustering dan formula dari rumus K Means Clustering. Pada metode Canopy K Means Clustering memiliki rumus formula yang berbeda dengan metode K Means Clustering pada beberapa tahapan pembelajaran. Perbedaan formula algoritma yaitu adanya penghitungan data dengan nilai jarak terdekat dan penghitungan data dengan jarak terjauh pada metode Canopy K Means Clustering. Persamaan antara kedua metode adalah konsep penghitungan jarak pada setiap iterasinya dengan menerapkan rumus pada penelitian ini dengan rumus Euclidean. Nilai yang diperoleh dipengaruhi oleh rumus pada setiap tahapan metode, dan metode dengan Canopy K Means Clustering cenderung menghasilkan nilai kesalahan yang lebih sedikit dibandingkan metode K Means Clustering hal ini menunjukkan bahwa korelasi angka yang dihasilkan oleh metode Canopy K Means Clustering lebih menghasilkan nilai yang sesuai kelompok data.

Metode K Means Clustering telah banyak diterapkan untuk pengelompokan data, dengan rumus pada metode K Means Clustering dimulai dengan inialisasi, penghitungan jarak dan penghitungan nilai cluster pada setiap iterasinya dengan metode jarak misalkan dengan Euclidean. Karakteristik nilai penghitungan kesalahan total diperoleh pada akhir iterasi setelah konvergen pada akhir iterasi. Pada penelitian [2] menerapkan metode K Means Clustering untuk mengelompokkan mahasiswa yang tidak lulus dan lulus pada jurusan manajemen berdasarkan beberapa variabel. Pada penelitian [3] K Means Clustering diterapkan sebagai pengelompokan produksi kelapa pada daerah Sumatra. Pada penelitian [4] menerapkan metode K Means Clustering sebagai sarana estimasi biaya yang dikeluarkan pasien. Pada penelitian [5] penerapan K Means Clustering sebagai pengelompokan minat beli produk konsumen. Pada penelitian [6] membahas pengembangan K Means Clustering pada pengolahan center. Pada penelitian [7] penerapan simulasi aplikasi K Means Clustering dengan sistem terdistribusi. Pada penelitian [12] pengembangan metode K Means Clustering pada penentuan nilai pada variabel K. Pada penelitian [15] pengembangan metode K Means Clustering pada center dengan penerapan konsep unsupervised pada beberapa variabel di center.

Metode Canopy K Means Clustering merupakan pengembangan dari metode K Means Clustering dan terdapat rumus yang berbeda pada saat penghitungan cluster pada setiap iterasinya sesuai dengan rumus pada metode Canopy K Means Clustering terdapat penghilangan data terdekat dan terjauh pada setiap iterasinya. Pada [1] menerapkan metode Canopy K Means Clustering pada pembagian *roadmap*. Pada penelitian [9] menerapkan metode Canopy K Means Clustering pada analisa edukasi siswa. Pada penelitian [10] penerapan image retrieval dengan dataset gambar yang digabungkan dengan Improved K Means Clustering. Pada penelitian [11] Pengaruh penerapan Algoritma Canopy pada metode K Means Clustering pada data dengan dimensi tinggi. Pada penelitian [13] adalah analisa metode canopy k means clustering dan hierarchy k means clustering. Pada penelitian [14] penerapan Canopy K Means Clustering pada data routing.

Pada penelitian ini menganalisa dengan menerapkan metode Canopy K Means Clustering dengan dataset gambar yang berbeda dan dibandingkan dengan metode K Means Clustering. Hasil analisa pengaruh formula jarak dengan rumus Euclidean diterapkan penghilangan data terjauh dan terdekat dengan metode Canopy K Means Clustering pada penelitian ini dengan data gambar yang terfilter menggunakan data set gambar yang diperoleh dari referensi data set dan selanjutnya difilter karena setiap produk memiliki karakteristik yang berbeda sebagai menambah atribut karakteristik data. Analisa kesalahan pada penelitian ini dengan menggunakan rumus dasar *Euclidean distance* yang menganalisa pengolahan cluster data pada metode K Means Clustering dan

Canopy K Means Clustering dimana terdapat pengolahan penghilangan data dengan jarak terjauh dan terdekat pada setiap iterasinya.

## 2. TINJAUAN TEORI

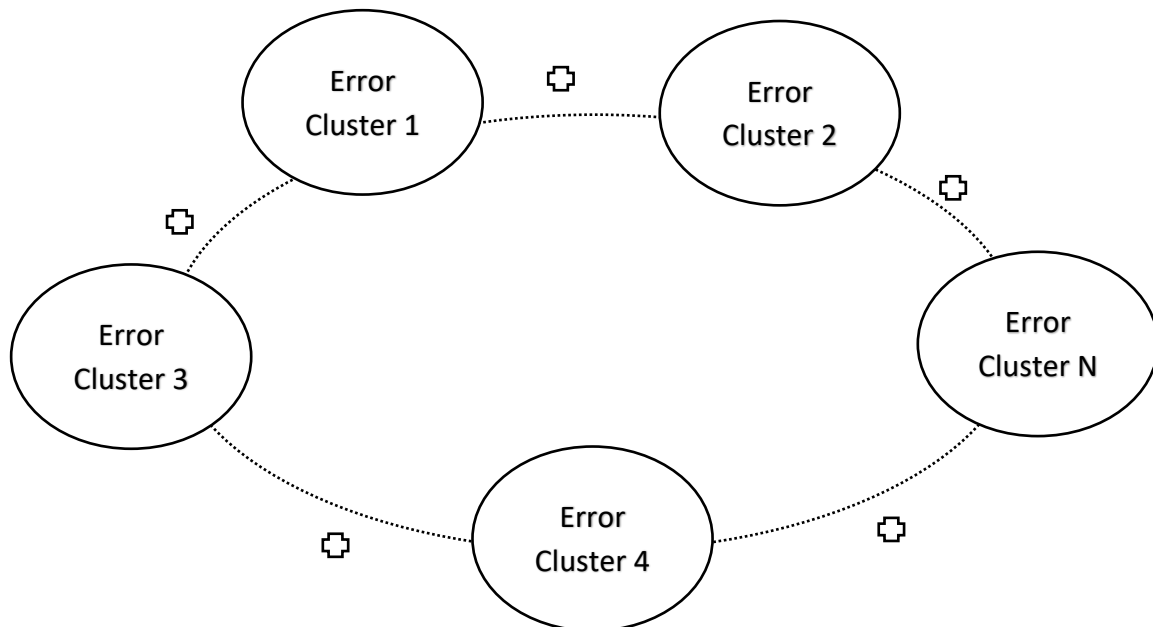
### 2.1 Metode Canopy K Means Clustering

Metode Canopy K Means Clustering adalah pengembangan dari metode K Means Clustering dengan perbedaan terdapat pada pengolahan nilai cluster pada setiap iterasinya. Algoritma metode Canopy K Means Clustering sebagaimana berikut [8]:

1. Inisialisasi nilai data yang akan diterapkan sebagai nilai awal cluster.
2. Proses penghitungan data pada setiap nilai cluster.
3. Penghitungan data dengan jarak terdekat.
4. Penghitungan data dengan jarak terjauh.
5. Penghitungan pembaharuan nilai cluster selain nilai terdekat dan terjauh.
6. Penghitungan nilai kesalahan pada setiap iterasi.
7. Berhenti iterasi jika iterasi lebih dari N iterasi jika tidak maka kembali ke no 2.

Data yang digunakan pada penelitian ini adalah data gambar yaitu data gambar berwarna yang diambil dari dataset dan selanjutnya difilter dengan menggunakan fungsi filter gambar yang terdapat pada program PHP.

Penghitungan kesalahan pada metode pengelompokan data dengan menerapkan rumus *Sum Square of Error* (SSE) yang menghitung kesalahan pada setiap kelompok dan dijumlah. Tujuan rumus SSE adalah menghitung kedekatan nilai kesalahan pada setiap iterasinya, dan tujuan dari pengolahan nilai *clustering* baik metode K Means Clustering ataupun metode Canopy K Means Clustering adalah memperbaiki nilai cluster agar nilai SSE semakin minimum.



Gambar 1. Konsep SSE pada Clustering

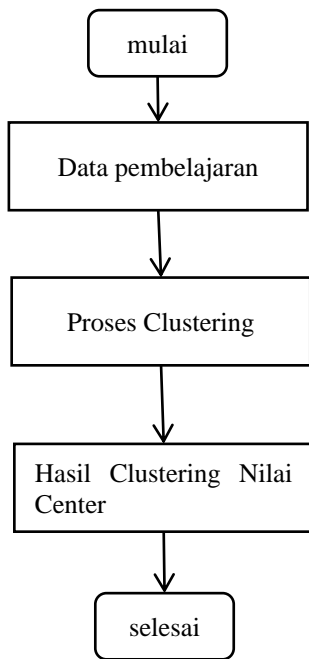
### 2.2 Filter Data

Filter data [16] adalah proses pengolahan gambar yang kompleks menjadi lebih sederhana misalkan dengan mengubah menjadi hitam dan putih, atau mengubah menjadi warna abu-abu ataupun menerapkan filter menjadi kabur dengan menerapkan filter rata-rata. Penerapan filter

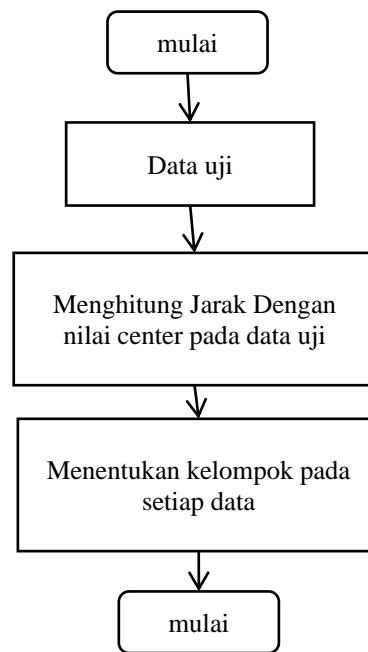
pada gambar diterapkan sebagai penyederhanaan gambar sebelum didata pada penghitungan dengan algoritma. Pada penelitian ini gambar dengan filter didata selanjutnya dihitung dengan algoritma K Means Clustering dan Canopy K Means Clustering. Filter yang diterapkan pada penelitian ini terdapat beberapa pilihan filter yang diterapkan dengan menggunakan fungsi dari bahasa pemrograman PHP.

### 3. METODE PENELITIAN

Terdapat beberapa tahapan penelitian pada penelitian ini yaitu membandingkan kedua metode. Metode yang dibandingkan adalah metode K Means Clustering dan Canopy K Means Clustering berdasarkan formula algoritma dari kedua metode.

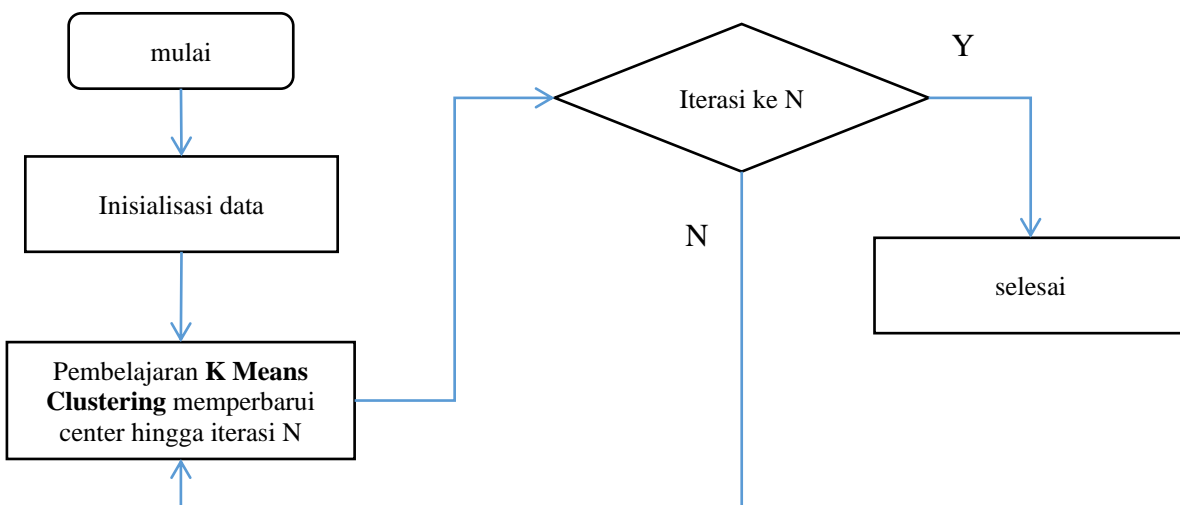


2.a Flowchart Pembelajaran Data

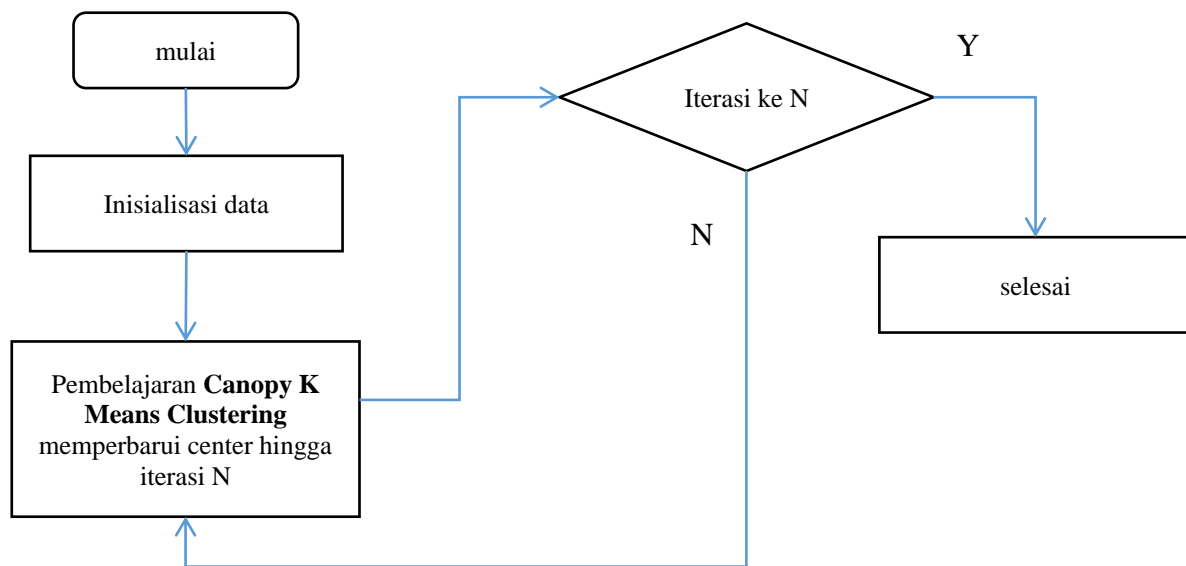


2.b Flowchart Uji Data

Gambar 2. Konsep Flowchart Pembelajaran dan Uji Data



Gambar 3. Flowchart Pembelajaran K Means Clustering

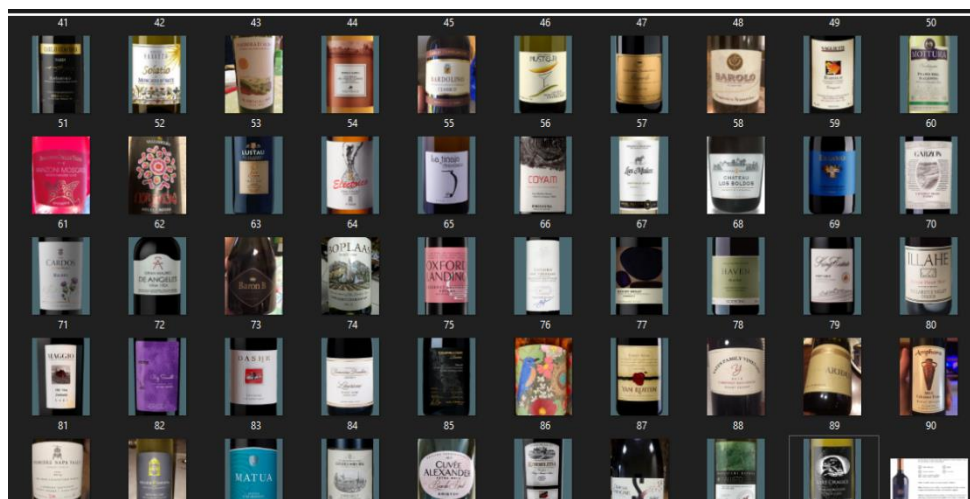


Gambar 4. Flowchart Pembelajaran Canopy K Means Clustering

Proses memperbarui data pada Gambar 3 pada metode K Means Clustering dengan menghitung nilai rata-rata data yang terkelompok pada setiap iterasinya yang dihitung dengan rumus *Euclidean distance* untuk pengelompokan datanya dan rumus rata-rata pada nilai center pada metode K Means Clustering. Pada Gambar 4 proses memperbarui nilai center pada metode Canopy K Means Clustering dengan menghitung rata-rata data yang terkelompok dengan mengurangi data yang memiliki nilai jarak terjauh sebagai perbedaan penghitungan nilai center dibandingkan dengan metode K Means Clustering dimana penghitungan jarak dengan metode *Euclidean distance*.

### 3.1 Metode Pengumpulan Data Gambar

Pengumpulan data gambar dengan mendownload dari dataset [17] dan disimpan di dalam database, data yang diakses dari dataset selanjutnya diproses dengan filter gambar. Jumlah data yang diuji yaitu 100 data gambar utama dan selanjutnya difilter sehingga total keseluruhan adalah 1000 data gambar.



Gambar 5. Beberapa Data Dataset Gambar[17].

### 3.2 Metode Pembangunan System

Pembangunan aplikasi dengan menerapkan bahasa pemrograman PHP dan basis data My SQL. Aplikasi yang dibangun pada penelitian ini terdapat dua metode cluster yaitu K Means Clustering dan Canopy K Means Clustering.

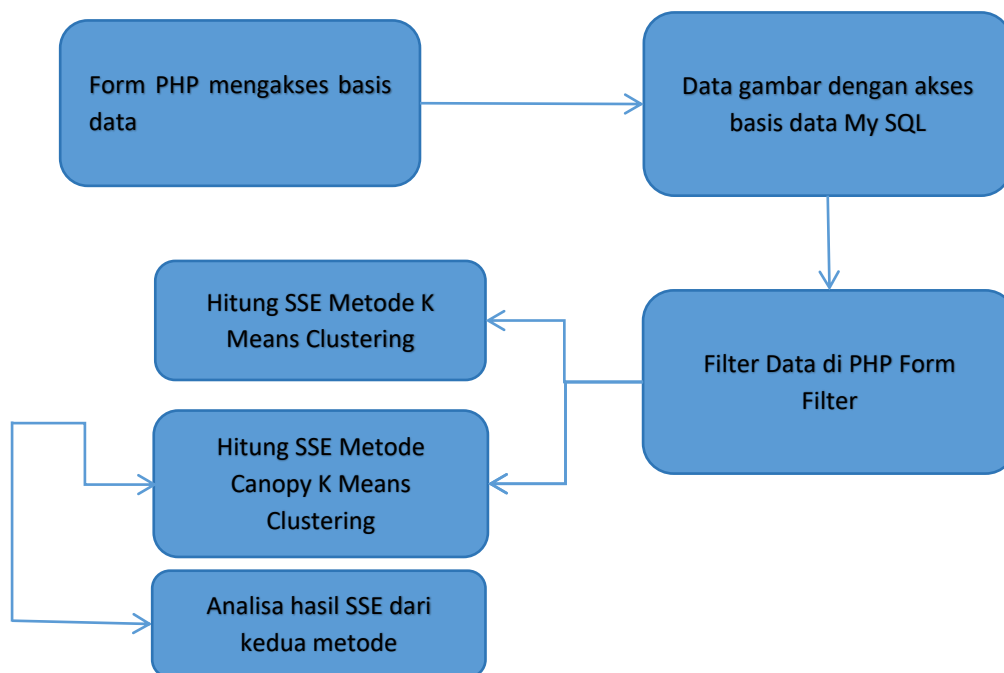
### 3.3 Metode Pengujian System

Pengujian pada penelitian ini dengan menerapkan data gambar yang terfilter sejumlah 1000 gambar dan dibelajarkan pada kedua metode dengan bahasa pemrograman PHP, dimana data dibagi menjadi N kali uji data pada setiap uji metode dari 100 data gambar utama. Jumlah uji sejumlah N kali uji data dengan data yang berbeda pada setiap kali uji coba dengan kombinasi dari 100 gambar.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Desain

Desain sistem yang diterapkan pada penelitian ini dengan menggunakan basis data My SQL dan program PHP, sebagaimana diagram pada gambar 6. Setiap data pada folder yang terdapat pada Gambar 2 akan melalui proses filter sejumlah 10 filter dengan menerapkan fungsi pada pemrograman PHP. Tujuan penerapan filter pada penelitian ini yaitu mengubah data bernilai 3 dimensi menjadi 2 dimensi sehingga mempermudah penghitungan nilai data pada algoritma K Means Clustering dan Canopy K Means Clustering.



Gambar 6. Analisa SSE pada kedua metode K Means Clustering dan Canopy K Means Clustering

### 4.2 Analisa SSE Pada K Means Clustering dan Canopy K Means Clustering

Sum Square of Error (SSE) pada metode clustering adalah perwujudan kedekatan hasil pengelompokan yang terbentuk dari hasil pembelajaran clustering. Sebagaimana Gambar 1, bahwa SSE dihitung pada setiap kesalahan masing-masing cluster, selanjutnya dijumlah pada seluruh cluster. Yang berpengaruh pada hasil angka pada SSE adalah data yang terlibat pada masing-masing metode, dimana dipengaruhi oleh rumus clustering. Pada penelitian ini penghitungan error dengan rumus penghitungan jarak yaitu metode *Euclidean distance*[2][3][4][5].

$$SSE = \sum_{i=1}^k \sum_{x_j \in C_i} |x_j - c_i| \quad (1)$$

Dimana k adalah jumlah kelompok atau cluster, dan x adalah jumlah seluruh data, misalkan terdapat 10000 data maka nilai x adalah antara 1 hingga 10000. Nilai i adalah jumlah kelompok, jika terdapat 3 kelompok maka nilai i antara 1 hingga 3.

Penghitungan SSE dipengaruhi oleh karakteristik data, jumlah data dan ukuran gambar yang dihitung, berikut contoh penghitungan SSE:

Jika terdapat 3 kelompok pada aplikasi, maka pada setiap iterasinya dihitung kesalahan pada masing-masing kelompok data.

Iterasi 1

Kesalahan hasil pada kelompok 1 dengan Euclidean adalah 5,67

Kesalahan hasil pada kelompok 2 dengan Euclidean adalah 3,2

Kesalahan hasil pada kelompok 3 dengan Euclidean adalah 2,5

Maka nilai SSE keseluruhan kelompok pada iterasi 1 adalah 5,67+3,2+2,5

Iterasi 2

Kesalahan hasil pada kelompok 1 dengan Euclidean adalah 5,65

Kesalahan hasil pada kelompok 2 dengan Euclidean adalah 3,1

Kesalahan hasil pada kelompok 3 dengan Euclidean adalah 2,3

Maka nilai SSE keseluruhan kelompok pada iterasi 2 adalah 5,65+3,1+2,3

Iterasi 3

Kesalahan hasil pada kelompok 1 dengan Euclidean adalah 4,7

Kesalahan hasil pada kelompok 2 dengan Euclidean adalah 2,8

Kesalahan hasil pada kelompok 3 dengan Euclidean adalah 2,1

Maka nilai SSE keseluruhan kelompok pada iterasi 3 adalah 4,7+2,8+2,1.

Hingga N iterasi, yang digunakan adalah iterasi terakhir yang paling minimum.

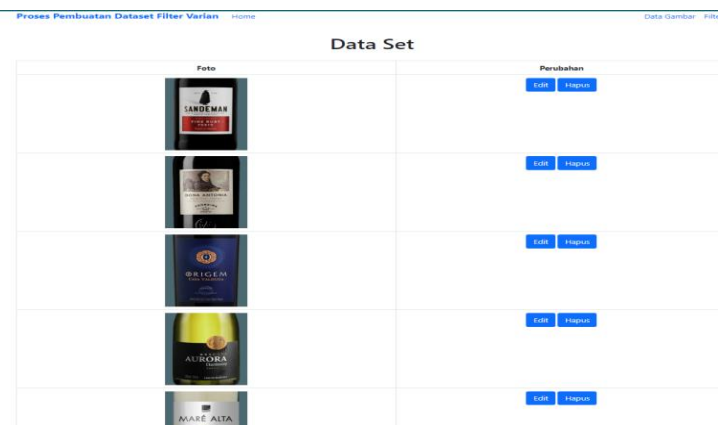
Perbedaan pengolahan data pada center dari kedua metode sebagaimana pada Tabel 1.

Tabel 1. Perbedaan pengolahan data dari dua Metode

K Means Clustering	Canopy K Means Clustering
Menggunakan seluruh data	Pada penghitungan setiap iterasi mengurangi data terdekat dan terjauh

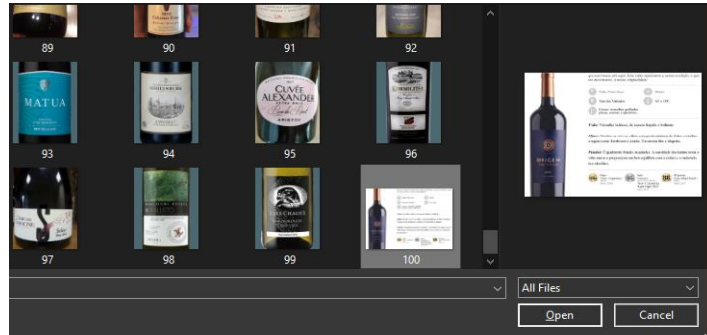
### 4.3 Hasil Uji

Pada penelitian ini terdapat 100 data utama dan pada Gambar 4 adalah uji pada penambahan data gambar.



Gambar 7. Data gambar yang digunakan pada aplikasi









Data ditambahkan di sistem dan disimpan pada folder, sesuai dengan kelas data dimana didalan folder akan didata keseluruhan filter.



Gambar 8. Penambahan Data gambar

Pada proses filter data akan diproses sebagaimana Gambar 9 dengan menggunakan fungsi yang terdapat pada PHP.

**Data Set**

Judul Buku	Foto Buku	Perubahan									
		Filter1	Filter2	Filter3	Filter4	Filter5	Filter6	Filter7	Filter8	Filter9	Filter10
		Filter1	Filter2	Filter3	Filter4	Filter5	Filter6	Filter7	Filter8	Filter9	Filter10
		Filter1	Filter2	Filter3	Filter4	Filter5	Filter6	Filter7	Filter8	Filter9	Filter10
		Filter1	Filter2	Filter3	Filter4	Filter5	Filter6	Filter7	Filter8	Filter9	Filter10
		Filter1	Filter2	Filter3	Filter4	Filter5	Filter6	Filter7	Filter8	Filter9	Filter10
		Filter1	Filter2	Filter3	Filter4	Filter5	Filter6	Filter7	Filter8	Filter9	Filter10
		Filter1	Filter2	Filter3	Filter4	Filter5	Filter6	Filter7	Filter8	Filter9	Filter10
		Filter1	Filter2	Filter3	Filter4	Filter5	Filter6	Filter7	Filter8	Filter9	Filter10

Gambar 9. Proses Filter Data gambar yang digunakan pada aplikasi

error iterasi ke3adalah9325.64705882

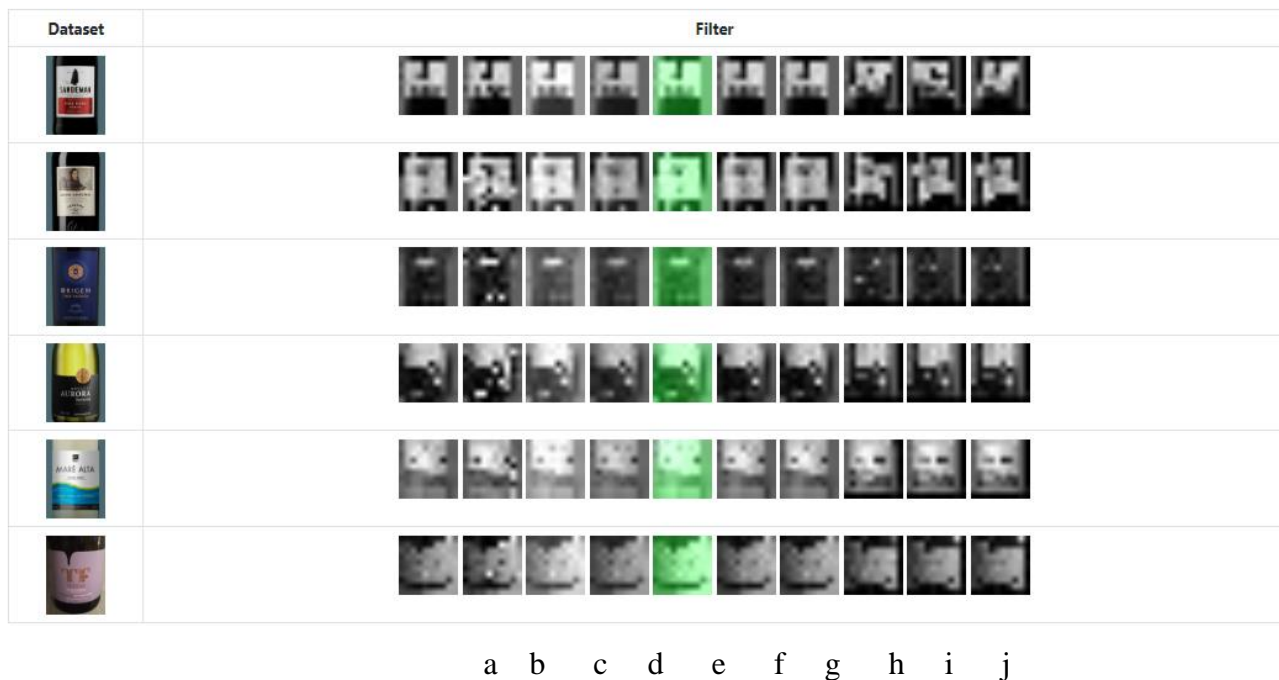
Prosentase After Canopy Clustering91.666666667%

Gambar 10. Uji Prosentase Hasil Pengelompokan

Prosentase hasil pengelompokan rata-rata 85%, pada Gambar 10 salah satu uji coba mendapatkan prosentase 91,6% keberhasilan kebenaran pengelompokan data, selain penghitungan SSE.



## Bangun Matrik Data Set



Gambar 11. Hasil Filter data Set

Hasil filter dengan fungsi grayscale

- a. Hasil filter dengan fungsi mean removal
- b. Hasil filter dengan fungsi brightness
- c. Hasil filter dengan fungsi contrast
- d. Hasil filter dengan fungsi colorize
- e. Hasil filter dengan fungsi gaussian blur
- f. Hasil filter dengan fungsi smooth
- g. Hasil filter dengan fungsi rotasi -5
- h. Hasil filter dengan fungsi rotasi 4
- i. Hasil filter dengan fungsi rotasi -4

Sebagaimana Gambar 11, gambar hasil filter melalui proses penyamaan atau keseragaman data yaitu normalisasi data sebelum dianalisa menggunakan kedua metode agar lebih sederhana pada kompleksitas data. Normalisasi dengan menyamakan ukuran data gambar yaitu pada aplikasi ini berukuran 10x10 piksel dengan jumlah data menjadi 1000 data gambar yang dibelajarkan dari 100 data awal yang belum difilter.

Basis data yang digunakan pada aplikasi ini adalah My SQL yang diterapkan sebagai sarana penyimpanan data dengan desain sebagaimana Gambar 12.

id	foto_produk	Filter1	Filter2	Filter3	Filter4	Filter5	Filter6	Filter7
4	4.jpeg	fotolestancoy/4/resizegray44.jpeg	fotolestancoy/4/resizeMEAN44.jpeg	fotolestancoy/4/resizeBRIGHT44.jpeg	fotolestancoy/4/resizeCONTRAST44.jpeg	fotolestancoy/4/resizeCOLORIZE44.jpeg	fotolestancoy/4/resizeGAUSSIAN44.jpeg	fotolestancoy/4/resizeSMOOTH
3	3.jpeg	fotolestancoy/3/resizegray33.jpeg	fotolestancoy/3/resizeMEAN33.jpeg	fotolestancoy/3/resizeBRIGHT33.jpeg	fotolestancoy/3/resizeCONTRAST33.jpeg	fotolestancoy/3/resizeCOLORIZE33.jpeg	fotolestancoy/3/resizeGAUSSIAN33.jpeg	fotolestancoy/3/resizeSMOOTH
1	1.jpeg	fotolestancoy/1/resizegray11.jpeg	fotolestancoy/1/resizeMEAN11.jpeg	fotolestancoy/1/resizeBRIGHT11.jpeg	fotolestancoy/1/resizeCONTRAST11.jpeg	fotolestancoy/1/resizeCOLORIZE11.jpeg	fotolestancoy/1/resizeGAUSSIAN11.jpeg	fotolestancoy/1/resizeSMOOTH
2	2.jpeg	fotolestancoy/2/resizegray22.jpeg	fotolestancoy/2/resizeMEAN22.jpeg	fotolestancoy/2/resizeBRIGHT22.jpeg	fotolestancoy/2/resizeCONTRAST22.jpeg	fotolestancoy/2/resizeCOLORIZE22.jpeg	fotolestancoy/2/resizeGAUSSIAN22.jpeg	fotolestancoy/2/resizeSMOOTH
5	5.jpeg	fotolestancoy/5/resizegray55.jpeg	fotolestancoy/5/resizeMEAN55.jpeg	fotolestancoy/5/resizeBRIGHT55.jpeg	fotolestancoy/5/resizeCONTRAST55.jpeg	fotolestancoy/5/resizeCOLORIZE55.jpeg	fotolestancoy/5/resizeGAUSSIAN55.jpeg	fotolestancoy/5/resizeSMOOTH
6	6.jpeg	fotolestancoy/6/resizegray66.jpeg	fotolestancoy/6/resizeMEAN66.jpeg	fotolestancoy/6/resizeBRIGHT66.jpeg	fotolestancoy/6/resizeCONTRAST66.jpeg	fotolestancoy/6/resizeCOLORIZE66.jpeg	fotolestancoy/6/resizeGAUSSIAN66.jpeg	fotolestancoy/6/resizeSMOOTH

Gambar 12. Penyimpanan data gambar filter



Gambar 13. Filter yang disimpan di folder

Tabel 2. Hasil Uji SSE Dataset Menggunakan K Means Clustering dan Canopy K Means Clustering

Uji Ke-	K Means Clustering	Canopy K Means Clustering
1	9325.64705882	9325,64705882
2	870977.823784	870831.318953
3	732.462745098	732.462745098

Berdasarkan hasil uji coba pada tabel 2, maka kecenderungan Algoritma Canopy K Means Clustering memiliki nilai SSE yang merupakan kedekatan data lebih baik dibandingkan dengan Algoritma K Means Clustering, dengan menggunakan beberapa dataset yaitu dataset wine, dan dataset cover buku. Jika dibandingkan dengan data penelitian sebelumnya, penerapan metode K Means Clustering dan Canopy K Means Clustering telah banyak diterapkan pada penelitian pengelompokan data dan kecenderungan data yang dikelompokkan dengan tipe data angka. Hasil pengelompokan clustering sangat dipengaruhi oleh data yang digunakan, dimana pada penelitian ini menerapkan data yang memiliki dimensi yang tinggi yaitu data gambar dengan jumlah dimensi yang telah disederhanakan.

## 5. KESIMPULAN

Sesuai dengan hasil uji coba dengan menggunakan data gambar dengan dataset yang berbeda yang memiliki dimensi yang tinggi pada setiap datanya adalah hasil uji coba menunjukkan bahwa pengaruh pengolahan data dengan algoritma Canopy K Means Clustering dapat mempengaruhi nilai SSE dimana pada hasil uji coba terdapat pengurangan yaitu 0,0264% dibandingkan algoritma K Means Clustering dengan rata-rata keberhasilan prosentase 85% pada hasil pengelompokan data. Nilai error dengan menggunakan metode Canopy K Means Clustering dipengaruhi oleh nilai center yang terbentuk dengan menggunakan rumus algoritma yang mengurangi data dengan jarak yang terjauh sehingga kecenderungan error yang dihasilkan pada

metode Canopy K Means Clustering lebih minimum jika dibandingkan dengan metode K Means Clustering.

## REFERENSI

- [1] X.Lin, J.Xu, "Road Network Partitioning Method Based on Canopy K- Means Clustering Algorithm", Archives of Transport, Volume 54 Issue 2, 95-106, 2020.
- [2] M. Darwis, L. H. Hasibuan, M. Firmansyah, N. Ahady, R. Tiaharyadini, "Implementation of K-Means Clustering Algorithm in Mapping the Groups of Graduated or Dropped-out Students in the Management Department of the National University", JISA (Jurnal Informatika dan Sains), Vol. 04, No. 01, June 2021.
- [3] M. Harahap, A. W. D. R. Zamili, M. A. Arvansyah, E. F. Saragih, S. Rajen, A. M. Husein, "K-Means Clustering Algorithm Approach in Clustering Data on Cocoa Production Results in the Sumatra Region", JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi) Vol. 6 No. 6 (2022) 905 - 910.
- [4] I. B. G. Sarasvananda, R. Wardoyo, A. K. Sari, "The K-Means Clustering Algorithm with Semantic Similarity to Estimate the Cost of Hospitalization", IJCCS (Indonesian Journal of Computing and Cybernetics Systems) Vol.13, No.4, October 2019, pp. 313-322.
- [5] I. A. Mulyo, J. Heikal, "Customer Clustering Using The K-Means Clustering Algorithm in Shopping Mall in Indonesia", Management Analysis Journal 365-371, ISSN 2252-6552.
- [6] H. Zhao, "Design and Implementation of an Improved K-Means Clustering Algorithm", Hindawi Mobile Information Systems Volume 2022, Article ID 6041484.
- [7] C. Wu, B. Yan, R. Yu, B. Yu, X. Zhou, Y. Yu, N. Chen, "k-Means Clustering Algorithm and Its Simulation Based on Distributed Computing Platform", Hindawi Complexity Volume 2021, Article ID 9446653.
- [8] N. S. Sagheer, S. A. Yousif, Canopy with k-means clustering algorithm for big data analytics, <https://doi.org/10.1063/5.0042398>, 2021.
- [9] R. Ananda, "Silhouette Density Canopy K-Means for Mapping the Quality of Education Based on the Results of the 2019 National Exam in Banyumas Regency", KHAZANAH INFORMATIKA, Vol. 5 No. 2, December 2019.
- [10] S. Sabena, Dr. P. Yogesh, L. SaiRamesh, Image Retrieval using Canopy and Improved K mean Clustering, International Conference on Emerging Technology Trends (ICETT) 2011.
- [11] S. Ambika, G. Kavitha, "Overcoming the Defects of K-Means Clustering by using Canopy Clustering Algorithm", IJSRD - International Journal for Scientific Research & Development | Vol. 4, Issue 05, 2016.
- [12] C. Yuan, H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm", MDPI Journal 2019, 2, 16; doi:10.3390/j2020016.
- [13] M. Alian, G. Al-Naymat, "Questions Clustering Using Canopy-K-means and Hierarchical-K means Clustering", Journal of Information Technology, and is available online at <https://doi.org/10.1007/s41870-022-01012-w> / .
- [14] M. Wu, Z. Li, J. Chen, Q. Min, T. Lu, A Dual Cluster-Head Energy-Efficient Routing Algorithm Based on Canopy Optimization and K-Means for WSN, Sensors 2022, 22, 9731.
- [15] K. P. Sinaga, M.S. Yang, Unsupervised K-Means Clustering Algorithm, Digital Object Identifier 10.1109/ACCESS.2020.2988796.
- [16] R.C. Gonzalez, R. E. Woods, Digital Image Processing Third Edition, Digital Image Processing Third Edition, Pearson Education, 2008.
- [17] <https://paperswithcode.com/dataset/x-wines>.