Jurnal Elektronik Ilmu Komputer Udayana p-ISSN: 2301-5373
Volume 11, No 2. November 2022 e-ISSN: 2654-5101

Perbandingan Pengelompokan Metode PSO K-Means Dan Tanpa PSO Dalam Pengelompokan Data Alert

I Gede Made Sankhya Saiyoga Krisna[1], I Wayan Supriana[2], I Dewa Made Bayu Atmaja Darmawan[3], Agus Muliantara[4], Ngurah Agus Sanjaya ER[5], Luh Gede Astuti[6]

[a]Informatics Engineering, Faculty of Math and Science, University of Udayana
South Kuta, Badung, Bali, Indonesia
[1]sankhyasaiyoga@gmail.com
[2]wayan.supriana@unud.ac.id
[3]dewabayu@unud.ac.id
[4]muliantara@unud.ac.id
[5]agus_sanjaya@unud.ac.id
[6]lg.astuti@unud.ac.id

Abstract

With increasing knowledge and increasing internet crime, an Intrusion Detection System (IDS) is needed, one of which is Snort which can detect attacks. An attack notification is needed to let administrators know if an attack has occurred. The grouping of alerts uses the PSO method on K-Means and continues with the calculation of the risk value to label the threat level, namely low, medium, high in each group. The Whatsapp bot will send groups of alerts that have high and medium labels only. A notification will appear on the Whatsapp application. The results obtained in this study by grouping the attack data, namely, the accuracy obtained by the system using the Particle Swarm Optimization method on K-Means obtained better results than only using the K-Means method.

Keyword**:** *Intrusion Detection System, K-Means, Snort*, *Clustering, log.*

1. Pendahuluan

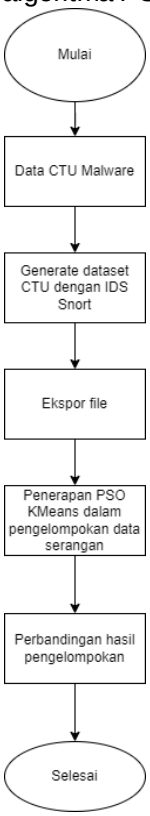
Untuk membantu seseorang administrator jaringan pada mengetahui jenis agresi yg berhasil lolos dalam sistem keamanan jaringan perlu dilakukan pengelompokan terhadap data log tadi sebagai akibatnya bisa dilakukan tindakan lebih lanjut dalam mengatasi agresi tadi oleh administrator jaringan. Tetapi, jumlah data dalam log Snort biasanya relatif *poly* sebagai akibatnya hal tadi bisa sebagai kasus lantaran buat menganalisisnya seseorang administrator jaringan membutuhkan *poly* waktu. Oleh karena itu, diperlukan suatu sistem yang dapat mengklasifikasikan data serangan menggunakan algoritma *KMeans* PSO. Tetapi agar mengetahui seberapa efektif optimasi penerapan algortima PSO pada *K-Means* perlu adanya perbandingan dengan menerapkan algoritma pengelompokan K-Means. Analisis *log* IDS Snort pernah dilakukan untuk seleksi notifikasi serangan dengan Algoritma K-Means sehingga hanya serangan berbahaya yang akan dikirimkan melalui SMS [1].

Berdasarkan pemaparan tersebut, penulis ingin mengelompokan jenis paket data *output Snort* menerapkan metode *K-Means* PSO setelah itu dibandingkan dengan hasil pengelompokan dengan algoritma *K-Means*. Hasil perbandingan tersebut berupa hasil nilai parameter *Silhouette, SSE,* dan kuantisasi. Data yang digunakan pada penelitian ini merupakan dataset yang berasal dari situs www.CTUMalware.com yang digunakan sebagai input *Snort*. Jumlah data pada penelitian ini adalah 1040 data dengan 9 atribut diantaranya *timestamp, protocol, sig id, sig rev*, *port* awal, *port* tujuan, *msg, src* dan *dst* .

2. Metode Penelitian

Penelitian ini menggunakan dataset yang diperoleh dari www.CTUMalware.com https://mcfp.felk.cvut.cz/publicDatasets/CTU-Malware-Capture-Botnet-54/, jumlah data pada penelitian ini adalah 1040 data dengan 9 atribut diantaranya *timestamp, protocol, sig id, sig rev*, *port* awal, *port* tujuan, *msg, src* dan *dst* . Penelitian ini dimulai dengan pengambilan data *snort Log,* data *snort log* merupakan hasil data dari penemuan yang dilakukan oleh *snort*, dan hasilnya berupa file *log* [1]. *Snort* menggunakan algoritma *KMeans* PSO dalam pengelompokan [2] untuk mengkonversi data ke format csv untuk kemudahan pemrosesan selama fase pengelompokan. Selain itu, data *log* yang dikonversi ke format csv melewati fase *preprocessing*. Pada fase *preprocessing* berguna untuk mengubah data menjadi format yang lebih mudah diproses dengan algoritma *KMeans* PSO [3]. Selain itu, setelah melalui tahap *preprocessing*, data tersebut dikelompokkan menggunakan algoritma PSO

KMeans, dan pengelompokan *KMeans* pada penelitian ini dilakukan dengan *Python* [4]. Tahap terakhir melakukan perbandingan antara penerapan algoritma PSO *Kmeans* dengan *Kmeans* [5].



Gambar 2. 1 System Flowchart

Langkah pertama dalam *preprocessing* adalah memproses nilai-nilai yang hilang. Nilai yang hilang adalah data yang tidak akurat karena informasi yang hilang membuat informasi yang terkandung di dalamnya tidak relevan. Kemudian, jika data tidak mengandung nilai yang hilang, langkah selanjutnya dalam preprocessing, data disesuaikan dengan rentang parameter dan dapat diproses oleh algoritma *Kmeans* PSO. Atribut yang digunakan pada penelitian ini adalah *port* awal*, port* tujuan, *msg*.

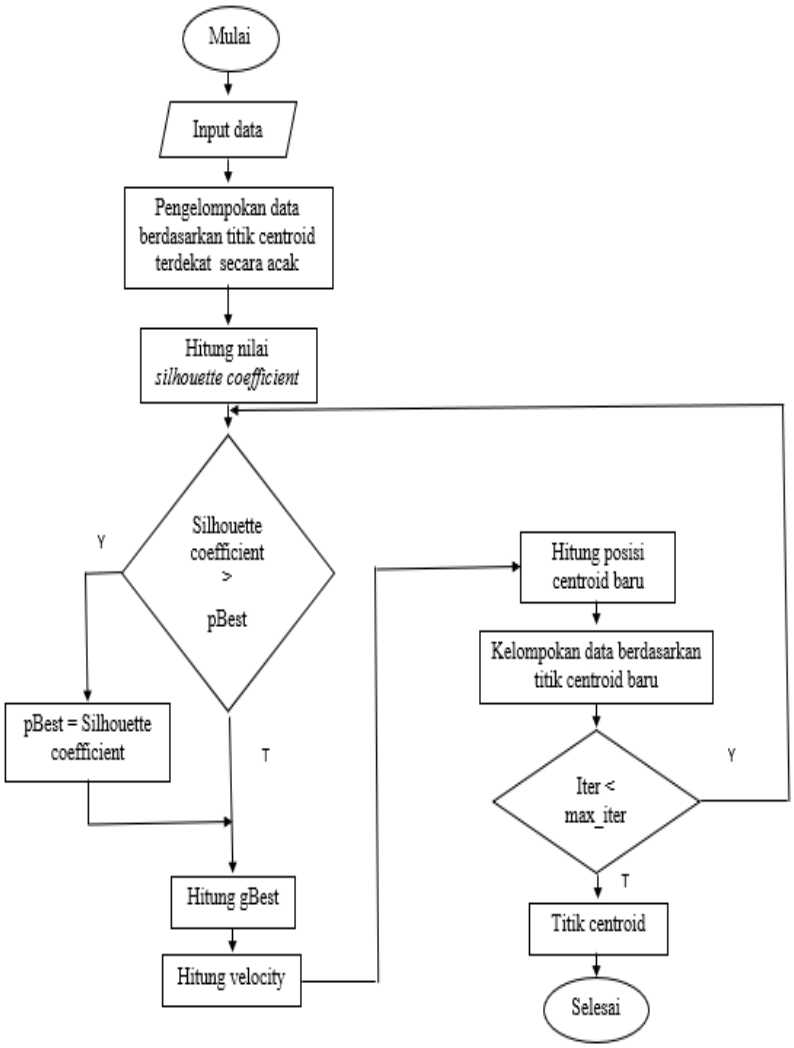
2.1 Clustering

Clustering adalah metode mengelompokkan data ke dalam *cluster*. Objek serupa berada di *cluster* yang sama, tetapi objek yang berbeda berada di *cluster* yang berbeda. Oleh karena itu, *clustering* adalah cara untuk mengelompokkan objek data ke dalam kelompok yang berbeda. B. Objek data yang sama yang masuk ke cluster yang sama dan objek data yang berbeda yang masuk ke *cluster* yang berbeda . Ada banyak metode pengelompokan, tergantung pada jenis data yang ingin Anda kelompokkan dan tujuan aplikasi Anda. Anda dapat menggunakan metode ini untuk mengelompokkan objek ke dalam kluster dan menggunakan hasil pengelompokan untuk mendeteksi keberadaan *outlier* dalam data Anda. Meskipun data yang digunakan adalah tipe data numerik.

2.2 PSO *K-Means Clustering*

Metode Particle Swarm Optimization (PSO) berfungsi dalam menentukan titik pusat tiap cluster atau centroid untuk digunakan dalam proses pengelompokan menggunakan algoritma K-Means. Algoritma K-Means dapat mengelompokkan kedalam beberapa cluster atau kelompok berdasarkan kemiripan dari data tersebut. Pada penelitian ini cluster yang akan dibentuk berjumlah 4 cluster/kelompok pada penilitian ini K-Means digunakan untuk mengelompokkan alert serangan yang tertangkap oleh Snort.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 2. *2* System Flowchart

Penjelasan flowchart :

1. Data serangan diambil dari database dari Snort yang selanjutnya akan dilakukan clustering.data yang digunakan untuk clustering yaitu data msg, jenis port dan data port tujuan.
2. Tentukan banyak cluster yang digunakan.
3. Menentukan nilai centroid awal. Nilai centroid awal ditentukan dengan menggunakan metode Particle Swarm Opitmizition (PSO).
4. Langkah pertama pada metode PSO yang dilakukan adalah menentukan titik centroid secara random serta mengelompokkan data pada titik centroid terdekat.
5. Kemudian dalam setiap iterasinya, dihitung fitness function menggunakan dari cluster yang terbentuk tersebut menggunakan silhouette coefficient.

$$s(i) = \frac{\min(d(i,j)}{\sum_{i,j}^{n}(d(i,j)} \quad (1)$$

Keterangan :
s = silhouette
d = distance
n = batas akhir
i,j = index

6. Setelah itu hitung kecepatan v menggunakan persamaan (2) dan hitung partikel x berdasarkan kecepatan menggunakan (3) . Iterasi berhenti apabila jumlah iterasi sudah melalui batas maksimal iterasi. Iterasi yang dilakukan penulis sebanyak 100 kali.

$$\mathrm{v}\frac{p}{i} = v\frac{p}{i} + c1.r1\left(localbest\frac{p}{i} - x\frac{p}{i}\right) + c2.r2\left(globalbest\frac{p}{i}\right) \quad (2)$$

Keterangan :
v = velocity
p = batas akhir
c1 = sosial komponen
c2 = kofnitif komponen
r1,r2 = random
i,j = index

7. Perbarui centroid

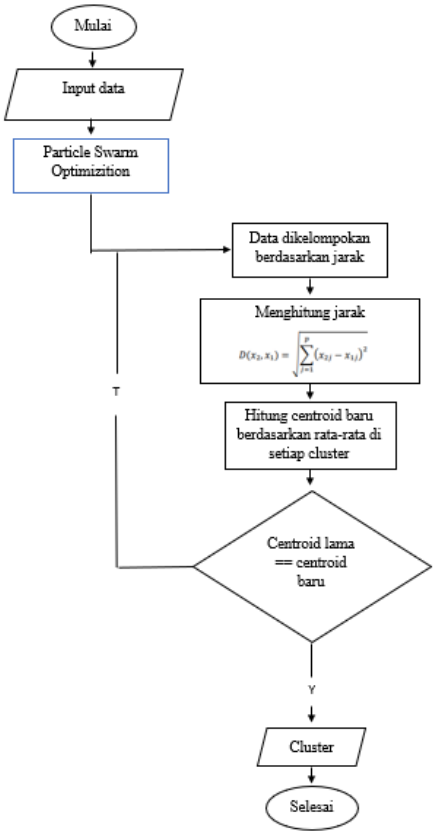
$$\mathrm{x}(i,j) = x(i,j + v(i,j) \quad (3)$$

Keterangan :
x = centroid
v = velocity

8. Hasil dari particle swarm optimization ini berupa titik centroid yang akan digunakan sebagai centroid awal dari k-means.

2.3 *K-Means Clustering*

Metode *Particle Swarm Optimization* (PSO) berfungsi pada memilih titik sentra tiap *cluster* atau *centroid* buat dipakai pada proses pengelompokan memakai prosedur pemecahan *K-Means*. Algoritma *K-Means* bisa mengelompokkan kedalam beberapa cluster atau gerombolan menurut kemiripan menurut data tersebut. Pada penelitian ini *cluster* yg akan dibuat berjumlah 4 cluster/gerombolan Nantinya setiap *cluster* akan melalui proses perhitungan nilai resiko terlebih dahulu buat menerima label masing-masing. Cluster yg berlabel low akan diabaikan sedangkan cluster yg berlabel high & medium akan dikirimkan menuju admin. Pada penilitian ini K-Means dipakai buat Langkah – Langkah Algoritma K-Means [6] :



Gambar 2. *3* System Flowchart

1. Tentukan K cluster centroid yang diinginkan.
2. Hitung jarak setiap data dengan centroid dengan Persamaan (1) dimana .Nx adalah jumlah data, Nf adalah jumlah dimensi, x_{ij} adalah data ke-I dengan atribut ke-j, dan c_{ij} adalah klaster ke-I dengan centroid ke-j.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

$$d(x_{ij}, c_{ij}) = \sqrt{\sum_{i=1}^{N_x} \sum_{j=1}^{N_f} (x_{ij} - c_{ij})^2} \quad (1)$$

3. Setiap data dikelompokan berdasarkan jarak terdekat dengan centroid
4. Perbarui centroid berdasarkan data yang telah dikelompokan dengan Persamaan (2) dimana x_j adalah atribut ke-j pada suatu data, N_d adalah jumlah dimensi, dan N_{xc} adalah jumlah data dalam satu klaster.

$$C_j = \frac{\sum_{j=1}^{N_d} x_j}{N_{xc}} \quad (2)$$

5. Ulangi hingga tidak terjadi perubahan pada nilai centroid, dan nilai jarak minimal maupun maksimal data ke cluster kurang dari threshold.

3. Hasil dan Pembahasan

Pada tahapan ini, penulis menentukan atibut yang digunakan pada proses pengelompokan selain itu penulis juga wajib menentukan nilai K. Atribut yang telah ditentukan yaitu *msg,* port tujuan dan port awal. Selanjutnya akan dilakukan perbandingan hasil parameter penerapan metode PSO *Kmeans* dengan tanpa PSO. Tahap awal yaitu dilakukan tahap *preprocessing.*

3.1 Preprocessing

Tahap *preprocessing* diawali dengan menangani data yang terdapat *missing value,* pada penelitian ini tidak terdapat *missing value.* Gambar 3.1 menunjukan bahwa data tidak terdapa *missing value.* Gambar 3.2 merupakan tahap *preprocessing* selanjutnya adalah data akan disesuaikan dengan rentang parameter sehingga dapat diproses dengan algoritma K-Means. pada tahap ini dilakukannya *scaling* data, scaling data berguna untuk perbandingan antara angka-angka agar membentuk nilai float dari angka 0 sampai 1.

Gambar 3. *1* Missing Value

```
df.isna().sum()

srcport    0
dstport    0
prio       0
pro        0
dtype: int64
```

Gambar 3. 2 Hasil Scaling

	srcport	dstport	prio
0	0.674054	0.752916	1.0
1	0.619336	1.000000	1.0
2	0.865373	0.408510	1.0
3	0.865373	0.408510	0.5
4	0.926963	0.159205	0.5

3.2 Clustering PSO *K-Means*

Pengujian tahap pertama yaitu implementasi PSO *K-Means* dalam pengelompokan data serangan. Pada Langkah pertama input jumlah kluster terlebih dahulu. Pada penelitian ini menginputkan jumlah cluster sebanyak 4. Langkah Selanjutnya jalankan program pada tabel 3.1 untuk melihat hasil dari proses clustering menggunakan algoritma PSO dan *K-Means*.

Setelah program dijalankan kemudian, dapat dilihat hasil dari proses PSO dan K-Means. Dimana tahap ini menghasilkan nilai dari Silhouette, SSE, dan Quantization. Pengujian ini dilakukan sebanyak 10 kali untuk dapat mencari rata -rata parameter hasil pengelompokan pada penerapan PSO pada K-Means.

Tabel 3. *1* Hasil Pengujian PSO *K-Means*

Pengujian	Silhouette	SSE	Quantization
Proses PSO K-Means	0.7491295267000625	13.60117717042267	1.570794247818463
	0.7491295267000728	13.57117717042267	1.580794247818463
	0.7491295267002971	13.57117717042266	1.580794247818462
	0.7131168439918436	16.31232372816058	1.590100594586362
	0.7491295267001408	13.57117717042266	1.580794247818462
	0.7440149394753982	13.91259238916166	1.549488707021564
	0.7491295266989284	13.57117717042268	1.580794247818463
	0.7478081172409544	16.07639176534623	1.696054258139580
	0.7491295258787264	13.57117717042267	1.580794247818463
	0.7431168439596913	13.37912960243446	1.589770950450701

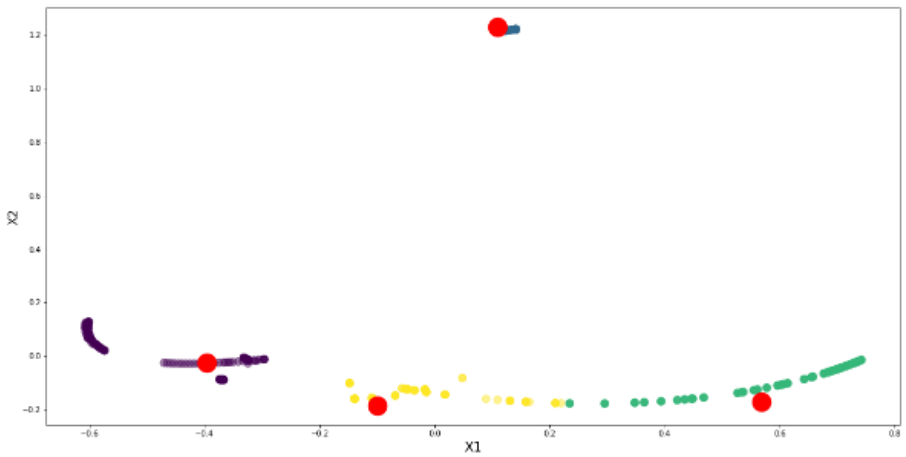
Rata – rata yang diperoleh setelah melakukan 10 pengujian, didapatkan hasil sebagai berikut.

Tabel 3. *2* Rata-rata Hasil Pengujian PSO *K-Means*

Silhouette	0.744283
SSE	14.11375
Quantization	1.590018

Gambar 3.3 merupakan tampilan hasil *cluster* dengan menerapkan algoritma PSO *K-Means.*

Gambar 3. 3 Gambar Hasil *Cluster* PSO *K-Means*



p-ISSN: 2301-5373
e-ISSN: 2654-5101

3. 3 **Clustering K-Means**

Pengujian tahap kedua yaitu implementasi *K-Means* dalam pengelompokan data serangan. Pada Langkah pertama input jumlah kluster terlebih dahulu. Pada penelitian ini menginputkan jumlah cluster sebanyak 4. Setelah program dijalankan kemudian, dapat dilihat hasil dari proses *K-Means*. Dimana tahap ini menghasilkan nilai dari *Silhouette, SSE*, dan *Quantization*. Pengujian ini dilakukan sebanyak 10 kali untuk dapat mencari rata -rata parameter hasil pengelompokan pada penerapan *K-Means*.

Tabel 3. *3* Hasil Pengujian *K-Means*

Pengujian	Silhouette	SSE	Quantization
Proses *K-Means*	0.7150454955578257	16.36147447336842	1.603585966152606
	0.7150454955567264	16.36147447336842	1.603585966152606
	0.7150454955577925	16.36147447336840	1.603585966152605
	0.7150454951273493	16.36147447336839	1.603585966152604
	0.7150454955283128	16.36147447336840	1.603585966152605
	0.7150454951268987	16.36147447336843	1.603585966152608
	0.7150454955556501	16.36147447336843	1.603585966152607
	0.7150454955551455	16.36147447336844	1.603585966152608
	0.7150454951265661	16.36147447336841	1.603585966152606
	0.7150454950951788	16.36147447336839	1.603585966152604

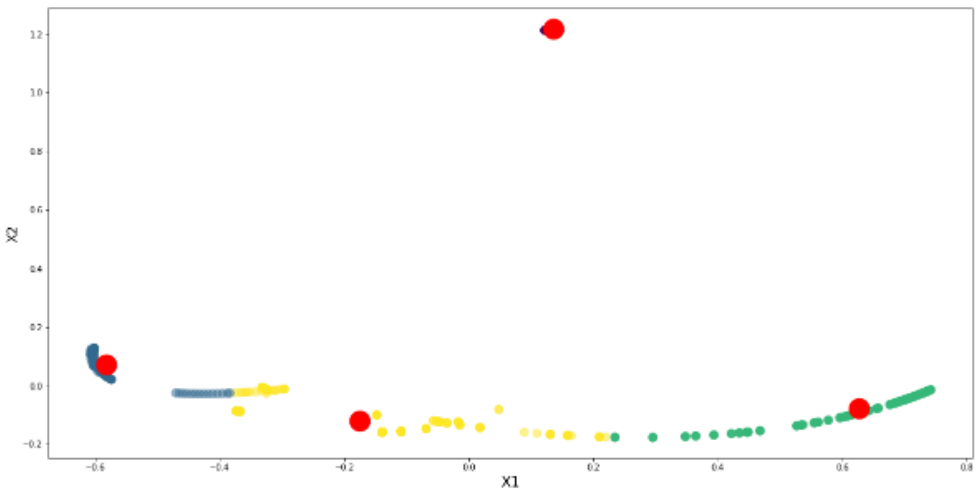
Rata – rata yang diperoleh setelah melakukan 10 pengujian, didapatkan hasil sebagai berikut.

Tabel 3. 4 Rata-rata Hasil Pengujian *K-Means*

Silhouette	0.715045
SSE	16.36147
Quantization	1.603585

Gambar 3. 4 Gambar Hasil *Cluster K-Means*

Gambar 3.3 merupakan tampilan hasil *cluster* dengan menerapkan algoritma PSO *K-Means.*



4. Kesimpulan

Setelah dilakukan perbandingan akurasi pada pengelompokan data serangan menggunakan metode PSO pada K-Means dan K-Means tentunya memperoleh hasil yang berbeda, dimana penerepan metode PSO pada K-Means menghasilkan nilai Sum of Square Error (SSE), Silhouette, dan Quantization error yang lebih baik dibandingkan hanya penerapan K-Means. Nilai Sum of Square Error (SSE), Silhouette, dan Quantization error pada metode optimasi PSO pada K-Means memperoleh hasil 14.11375, 0.744283, dan 1.590018. Sedangkan nilai Nilai Sum of Square Error (SSE), Silhouette, dan Quantization error pada metode K-Means memperoleh nilai 16.36147, 0.715045, dan 1.590018.

Referensi

[1] B. Alfiansyah, S. Syaifuddin, and D. Risqiwati, "Pengelompokan Notifikasi Alert Intrusion Detection System Snort Pada Bot Telegram Menggunakan Algoritma K-Means," *J. Repos.*, vol. 2, no. 3, p. 339, 2020, doi: 10.22219/repositor.v2i3.436.

[2] A. Y. Ananta, "Seleksi Notifikasi Serangan Berbasis Ids Snort Menggunakan Metode K-Means," *SMARTICS J.*, vol. 3, no. 2, pp. 31–37, 2017, doi: 10.21067/smartics.v3i2.1954.

[3] M. Affandi *et al.*, "Implementasi Snort Sebagai Alat Pendeteksi Intrusi Menggunakan Linux," *J. Teknol. Inf.*, vol. 4, no. 2, 2013, [Online]. Available: www.linux.org.

[4] M. Khalid, N. Pal, and K. Arora, "Clustering of Image Data Using K-Means and Fuzzy K-Means," *Int. J. Adv. Comput. Sci. Appl.*, vol. 5, no. 7, pp. 160–163, 2014, doi: 10.14569/ijacsa.2014.050724.

[5] M. A. Ramadhan, Efendi, "Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling," *Semin. Nas. Teknol. Informasi, Komun. dan Ind. 9*, pp. 219–226, 2017.

[6] F. Y. Bisilisin, Y. Herdiyeni, and B. P. Silalahi, "Optimasi K-Means Clustering Menggunakan Particle Swarm Optimization pada Sistem Identifikasi Tumbuhan Obat Berbasis Citra," *J. Ilmu Komput. dan Agri-Informatika*, vol. 3, no. 1, p. 37, 2017, doi: 10.29244/jika.3.1.37-46.