

Klasterisasi Kualitas Minyak Kelapa Sawit Menggunakan Metode K-Means Clustering (Studi kasus PT Tasma Puja)

Antoni Pribadi¹, Andri Nofiar, Am²

^{1,2} Teknik Informatika Politeknik Kampar

Jln. Tengku Muhammad KM 2 Bangkinang INDONESIA

[1antonipribadi.polkam@gmail.com](mailto:antonipribadi.polkam@gmail.com), [2andrinofiar90@gmail.com](mailto:andrinofiar90@gmail.com)

Intisari— Klasterisasi kualitas minyak kelapa sawit pada PT. Tasma Puja dilakukan dengan cara uji laboratorium dan selanjutnya data disimpan menggunakan *database* menggunakan *microsoft excel*. Klasterisasi data tersebut membutuhkan waktu dan memungkinkan data hilang, kedepannya perkembangan ilmu pengetahuan nantinya bisa di gantikan dengan pendekatan *data mining* yang dapat digunakan dalam mengelompokkan kualitas minyak kelapa sawit berdasarkan standarnya. Kriteria untuk menentukan mutu minyak kelapa sawit diambil dari data kandungan air, kotoran dan asam lemak bebas. Dari data yang didapat digunakanlah *metode k-means clustering* untuk menentukan kualitas minyak kelapa sawit. Metode *k-means clustering* dapat diterapkan untuk mengelompokkan suatu kualitas minyak kelapa sawit berdasarkan kriterianya. Data yang digunakan adalah data kualitas minyak kelapa sawit pada bulan desember 2017. Hasil pengujian menggunakan *software RapidMiner* terdapat 3 *cluster* yaitu *cluster 0* kategori baik berjumlah 12 data, *cluster 1* kategori sangat baik berjumlah 13 data dan *cluster 2* kategori kurang baik berjumlah 6 data. Metode *k-means clustering* dapat digunakan untuk proses pengolahan data menggunakan konsep *data mining* dalam mengelompokkan data sesuai kriterianya.

Keywords— Kalasterisasi, Data mining, K-means clustering, RapidMiner

Abstract— *Clustering the quality of palm oil at PT. Tasma Puja is done by laboratory testing and then the data is stored using a database using Microsoft Excel. Clustering the data requires time and allows data to be lost, in the future the development of science can later be replaced by a data mining approach that can be used in classifying the quality of palm oil based on its standards. Criteria for determining the quality of palm oil are taken from data on water content, impurities and free fatty acids. From the data obtained, the k-means clustering method is used to determine the quality of palm oil. The k-means clustering method can be applied to classify a quality of palm oil based on the criteria. The data used are data on the quality of palm oil in December 2017. The test results using the RapidMiner software there are 3 clusters namely cluster 0 good categories totaling 12 data, cluster 1 very good categories totaling 13 data and cluster 2 categories not good amounting to 6 data. The k-means clustering method can be used for data processing using the concept of data mining in grouping data according to its criteria.*

Keywords— Kalasterisasi, Data mining, K-means clustering, RapidMiner

I. PENDAHULUAN

Berdasarkan Badan Pusat Statistik (BPS) jumlah total luas area perkebunan kelapa sawit di Indonesia pada saat ini mencapai 11,9 juta hektar. Perkebunan kelapa sawit banyak tersebar di berbagai daerah, seperti Sumatera, Jawa, Kalimantan, dan Sulawesi. Riau merupakan provinsi yang banyak memiliki perkebunan kelapa sawit. Industri yang mengelola kelapa sawit salah satunya yaitu PT. Tasma Puja yang beralamat Jl. Sungai Kuamang, Kampar, dimana ia memproduksi minyak kasar atau minyak mentah dari minyak makan dari nabati dan hewani.

Untuk menghitung kualitas minyak kelapa sawit di PT. Tasma Puja dilakukan uji laboratorium berdasarkan kandungan air, kotoran dan asam lemak bebas. Berdasarkan pada bobot yang telah ditetapkan baik secara nasional maupun internasional, dampak yang sangat mempengaruhi kualitas minyak kelapa sawit terutama terletak pada

kandungan asam lemak bebas (ALB). Kandungan asam lemak bebas yang baik sebesar 3,0% [1].

Proses penyimpanan data air, kotoran dan asam lemak bebas yang dihasilkan setiap kali produksi dari hasil uji laboratorium masih disimpan secara manual menggunakan aplikasi *Microsoft Excel*. Dengan penyimpanan data tersebut terdapat kelemahan yaitu sulitnya dalam melakukan penilaian kualitas produksi minyak kelapa sawit yang dihasilkan. Produksi minyak kelapa sawit yang dilakukan setiap hari pada PT. Tasma Puja menyebabkan banyaknya penyimpanan data yang berdampak pada lamanya waktu yang dibutuhkan untuk mencari hasil terbaik dari setiap kali produksi minyak kelapa sawit.

Dengan perkembangan ilmu pengetahuan yang semakin pesat nantinya, proses penyimpanan data hasil produksi pada PT. Tasma Puja dapat digantikan dengan pendekatan *data mining* yang digunakan dalam

mengelompokkan kualitas minyak kelapa sawit berdasarkan bobot yang telah ditetapkan, dengan melakukan pembagian data kedalam kelas atau *cluster* berdasarkan tingkat kesamaannya. Sehingga Data yang diolah dengan menggunakan *data mining* menghasilkan pengetahuan atau *Knowledge* yang baru, dengan bersumber dari data lama hasil produksi minyak kelapa sawit.

Knowledge Discovery in Database (KDD) merupakan suatu metode yang digunakan untuk menganalisis data pada penerapan *data mining* untuk mencari pengetahuan, dan informasi yang lebih dari sebuah kumpulan data. *Knowledge Discovery in Database* (KDD) yaitu suatu kegiatan yang terdiri dari beberapa tahapan, seperti *Cleaning and Integration, Selection and Transformation, Data mining, dan Evaluation and Interpretation*. Pada proses *Evaluation and Interpretation* ini merupakan cakupan dari proses pemeriksaan baik pola itu atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa sebelumnya. Dalam memilih suatu metode atau *algoritma* yang digunakan sangat mempengaruhi dari proses dan tujuan dari *Knowledge Discovery in Database* (KDD)[2].

Data mining merupakan serangkaian proses untuk mendapatkan informasi yang berharga dari data, dimana suatu proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data dengan ukuran yang cukup besar. *Data mining* dapat dilakukan berbagai bidang dengan mempunyai sejumlah data. *Data mining* ini juga dapat disebut campuran dari statistik, kecerdasan buatan dan riset basis data yang masih berkembang[3].

Clustering merupakan salah satu teknik pengelompokkan dalam data mining yaitu pengelompokkan sejumlah data ataupun objek ke dalam suatu kelompok atau *cluster* atau *group* sehingga setiap data dalam *cluster* tersebut berisi suatu data yang memiliki kemiripan ataupun berbeda dengan objek dalam *cluster* yang lain[4].

Algoritma K-Means merupakan suatu metode yang digunakan untuk menemukan kelompok dari dokumen atau objek yang *non-overlapping*, dan juga dapat disebut sebagai suatu *algoritma* yang paling efektif untuk mengelompokkan kumpulan data. *Algoritma* ini dapat diimplementasikan, dijalankan dan diproses dengan cepat, mudah beradaptasi[5].

Adapun langkah-langkah perhitungan menggunakan algoritma K-Means ini yaitu :

1. Masukkan jumlah *K cluster*, *K* disini berarti banyaknya *cluster* yang ingin dibentuk.
2. Tentukan nilai awal *centroid* (*K* pusat *cluster*) untuk *K cluster* dari sejumlah dataset yang bisa dilakukan dengan berbagai cara yang salah satunya dengan secara acak atau *random*.
3. Semua data/objek dialokasikan ke *cluster* yang terdekat antara dua objek yang sesuai dengan jarak kedua objek tersebut. Sama halnya dengan kedekatan satu data ke *cluster* tertentu. Langkah ini dapat dihitung dari jarak tiap data ke tiap pusat *cluster* (*centroid cluster*). Untuk menghitung setiap titik pada *cluster* terdekat bisa menggunakan rumus jarak *Euclidean*, sebagai berikut :

$$D_{(i,j)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

Dimana:

$D_{(i,j)}$ = jarak data ke *i* ke pusat *cluster j*

X_{ki} = Data ke *i* pada atribut data ke *k*

X_{kj} = Titik pusat ke *j* pada atribut ke *k*

4. Hitung ulang untuk mencari nilai *centroid* dari setiap *K cluster* dengan jumlah data *m* untuk menemukan nilai *centroid K cluster* yang baru. Adapun titik pusat *cluster* (*centroid cluster*) yaitu rata-rata dari semua titik dalam *cluster*. Adapun cara mencarinya bisa menggunakan *median* dari data kelompok tersebut[6].

II. METODE PENELITIAN

Metode penelitian merupakan kaidah yang harus dilakukan dalam sebuah penelitian. Secara umum metodologi yang artinya metode melakukan sesuatu memakai pikiran secara teliti demi mencapai tujuan sedangkan arti dari penelitian adalah kegiatan mencatat, mencari, merumuskan dan menganalisis sampai penyusunan pelaporan. Pada metodologi penelitian ini menjelaskan tentang bagaimana kerangka kerja penelitian yang akan dilaksanakan, hal ini dimaksudkan agar hasil yang diperoleh dari penelitian lebih maksimal.

Kerangka Kerja

langkah-langkah kerja penelitian sebagai berikut :

1. Mengidentifikasi Masalah
Masalah yang akan diteliti harus ditentukan terlebih dahulu, karena tanpa mampu menentukan dan mendefinisikan batasan masalah yang akan diteliti, maka tidak akan pernah dapat suatu solusi yang terbaik dari masalah tersebut. Masalah yang diidentifikasi dalam penelitian ini adalah mengelompokkan asam lemak bebas untuk menentukan kualitas minyak kelapa sawit yang dihasilkan pada PT. Tasma Puja dengan menggunakan *software RipedMiner*.
2. Menganalisis Masalah
Analisis masalah adalah untuk dapat memahami masalah yang telah ditentukan ruang lingkup dan batasannya. Dengan menganalisa masalah yang telah ditentukan tersebut, maka diharapkan masalah dapat dipahami dengan baik. Proses pengambilan keputusan dalam pengelompokan kualitas minyak kelapa sawit pada PT. Tasma Puja.
3. Mengumpulkan Data
Untuk memperoleh data-data yang dibutuhkan dalam proses penulisan tesis ini, maka penulis menggunakan metode-metode sebagai berikut :
 - a. Studi Kepustakaan (*Library Research*)
Studi kepustakaan adalah suatu metode

pengumpulan data yang dilakukan pada perpustakaan yang harus diperhatikan sumber yang diteliti yaitu buku bacaan yang menjadi pembahasan sumber masalah yang dibahas, dan bahan-bahan yang berhubungan dengan masalah yang diteliti.

- b. Pengamatan (*Observation*)
Untuk mengumpulkan informasi dilakukan pengamatan langsung pada laboratorium yang bertujuan untuk mendapatkan keterangan yang berhubungan dengan objek yang akan diteliti.
- c. Dokumentasi
Dokumentasi merupakan cara memperoleh data dan informasi yang dibutuhkan dengan mengumpulkan data dari dokumen perusahaan yang berkaitan dengan proses pengumpulan data.

4. Mempelajari Literatur

Setelah mengumpulkan data yang akan di proses dalam studi literatur ini, terlebih dahulu mengetahui pengetahuan yang akan diterapkan dalam metode ini, literatur yang akan dipelajari ini bersumber dari jurnal-jurnal ilmiah yang di publikasikan di internet.

5. Memproses Data ke Algoritma *K-Means Clustering*

Setelah melakukan pengumpulan data dan mempelajari literatur, selanjutnya dilakukan proses ke algoritma *K-Means Clustering* yaitu :

1. Menentukan jumlah *cluster*
2. Menentukan titik pusat *centroid* secara *random*
3. Menghitung jarak terdekat *centroid*
4. Mengelompokkan data ke masing-masing *cluster*

Maka didapatkanlah pengetahuan atau *knowledge* baru yang bersumber dari data lama.

6. Menguji dan Analisa Hasil

Penulis melakukan pelatihan dan pengujian setelah melakukan pengumpulan data, proses ke algoritma *K-Means Clustering*, dan penginstalan *software* yang akan digunakan yaitu *RipidMiner*, untuk selanjutnya melakukan pelatihan dan pengujian terhadap data. Pengujian adalah data yang diperoleh dari pemrosesan data awal dengan menggunakan metode *K-Means Clustering*. Hasil yang diperoleh dari pengujian kemudian akan dianalisa agar kondisi lebih akurat.

7. Hasil dan Pembahasan

Kesimpulan hasil dari pengelompokan kualitas minyak kelapa sawit berdasarkan air,

kotoran dan asam lemak bebas menjadi baik, sangat baik dan kurang baik, dengan menggunakan metode *K-Means Clustering* dari tahapan proses menguji dan analisa hasil, sehingga menghasilkan kesimpulan yang baik.

III. Hasil dan Pembahasan

Hasil dan pembahasan pada penelitian ini meliputi : proses seleksi data, proses pengolahan data dan hasil pengolahan data.

Proses Seleksi Data

Dalam proses *Knowledge Discovery in Database* (KDD) seleksi data merupakan proses himpunan data dan menciptakan data target pada sampel data, dimana penemuan akan dilakukan selanjutnya hasil seleksi disimpan dalam suatu berkas yang terpisah dari data basis data operasional. Setelah data didapatkan dari pihak PT. Tasma Puja seperti data awal diatas maka dilakukan proses seleksi data, proses ini dilakukan agar memudahkan proses perhitungan algoritma *K-Means Clustering*. Adapun data yang digunakan adalah data pada bulan Desember 2017 yang dapat dilihat pada Tabel berikut :

Tabel 1. Daftar seleksi data bulan desember

NO	Air	Kotoran	ALB
1	0.280	0.019	3.610
2	0.000	0.000	0.000
3	0.000	0.000	0.000
4	0.320	0.019	3.520
5	0.310	0.019	3.750
6	0.260	0.019	3.200
7	0.340	0.019	3.370
8	0.320	0.019	3.510
9	0.290	0.010	3.840
10	0.000	0.000	0.000
11	0.230	0.019	4.340
12	0.350	0.019	4.040
13	0.350	0.018	3.810
14	0.310	0.018	3.600
15	0.450	0.019	3.990
16	0.340	0.019	3.710
17	0.000	0.000	0.000
18	0.280	0.019	3.560
19	0.290	0.020	4.160
20	0.240	0.019	3.460
21	0.300	0.019	3.940
22	0.310	0.019	3.960
23	0.270	0.019	4.120
24	0.270	0.019	4.310
25	0.000	0.000	0.000
26	0.240	0.019	4.790
27	0.300	0.019	4.190
28	0.270	0.019	4.080
29	0.300	0.018	4.540
30	0.320	0.018	4.310
31	0.000	0.000	0.000

Proses Pengolahan Data

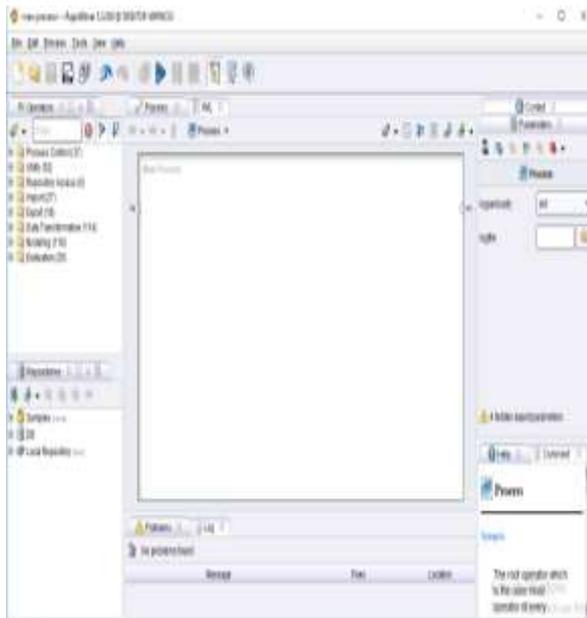
Proses pengolahan data menggunakan *software RapidMiner* memiliki Langkah-langkah sebagai berikut :

1. Langkah ke 1, membuka *software RapidMiner* kemudian pilih *New Process* seperti gambar 1 berikut :



Gambar 1. Tampilan Halaman Awal *Rapidminer*

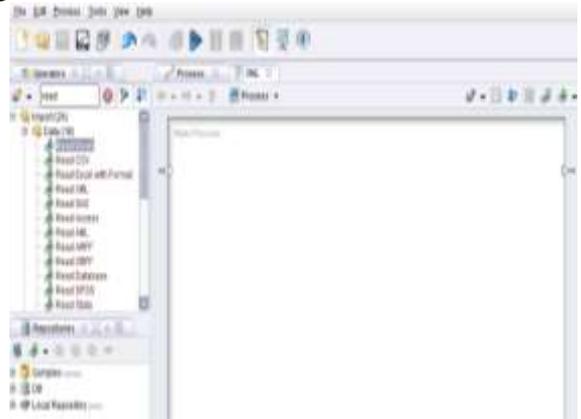
2. Langkah ke 2, membuat *New Process* dimana akan tampil *Window Design Perspectives* yang merupakan area kerja pada *software Rapidminer*, seperti gambar berikut :



Gambar 2. Tampilan Halaman *Window Design Perspectives*

3. Langkah ke 3, mengimport data yang dapat dibaca langsung diantaranya berupa *file* dengan ekstensi *.xls* atau *.xlsx*, yakni *file* dari *Ms. Excel*. Caranya dengan

mengklik kanan pada *mouse* di layar *process*, *insert operator*, *data access*, *files*, *read*, *read excel*, seperti gambar berikut :



Gambar 3. Tampilan Halaman *Import Data*

4. Langkah ke 4, merupakan cara mengimport *data*, pada tahap ini ada dua cara bisa klik *Add Data* atau klik *import configuration wizard*, seperti gambar berikut :



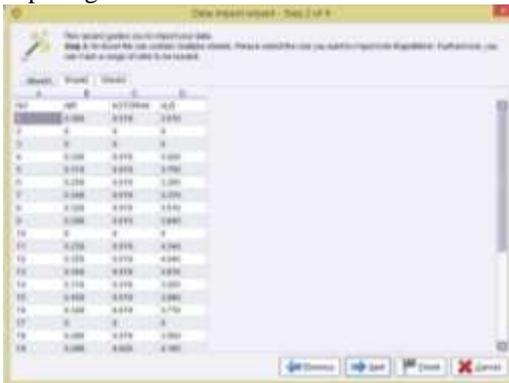
Gambar 4. Tampilan Halaman Cara *Import Data*

5. Langkah ke 5, dengan memilih *import configuration wizard* sesuai Gambar 5.5 diatas, selanjutnya memilih dimana tempat penyimpanan *data*, setelah *data* dipilih kemudian klik tombol *next*. Seperti gambar berikut :



Gambar 5. Tampilan Halaman Pilih *Data Yang Di-Import*

6. Langkah ke 6, selanjutnya akan tampil *data import wizard* maka akan muncul data *excel* yang di-import tadi seperti gambar berikut :



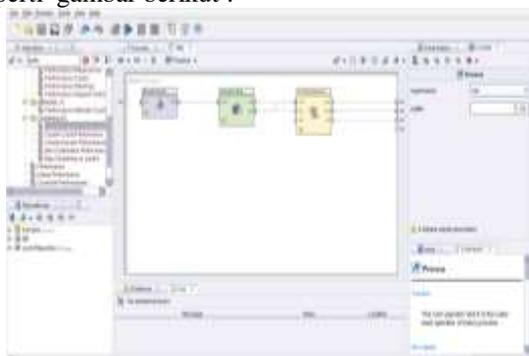
Gambar 6. Tampilan Halaman *Data Import Wizard*

7. Langkah ke 7, setelah setiap tahapan diatas di lakukan dengan melanjutkan dengan klik tombol *next*, selanjutnya mengganti attribute pada NO menjadi id dan klik tombol *finish* seperti gambar berikut :



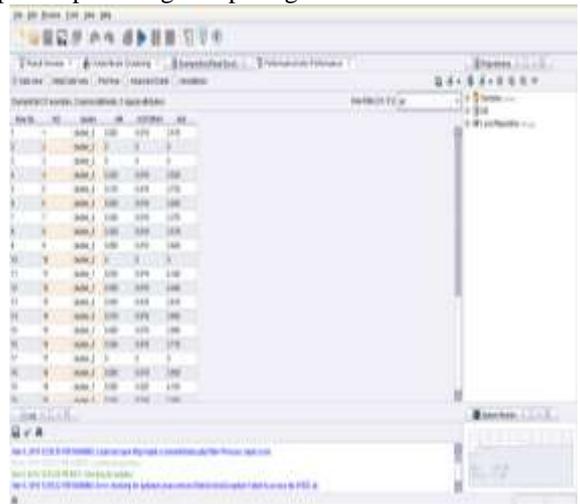
Gambar 7. Tampilan Halaman *Data Pada Import Wizard*

8. Langkah ke 8, Selanjutnya lakukan langkah-langkah yaitu pada operator *views*, pilih folder *Modelling* > *Segmentation* > *K-Means*. Drag and drop *K-Means* ke *Main Process*, tentukan jumlah cluster dan hubungkan *output* pada *repository* ke arah *clustering* kemudian *clustering* ke arah *result*. Terakhir klik Tombol *Play*, seperti gambar berikut :



Gambar 8. Tampilan Halaman Pengaturan *K-Means*

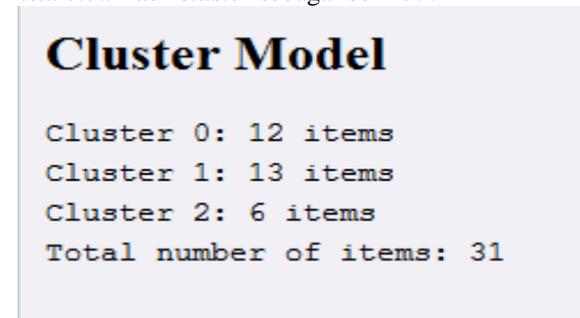
9. Langkah ke 9, Setelah proses dijalankan akan tampil *window results perspectives* sebagai halaman awal hasil proses perhitungan seperti gambar berikut :



Gambar 9. Tampilan Halaman Awal Hasil Proses Perhitungan

Hasil Pengolahan Data

Hasil dari eksekusi pengolahan data kualitas minyak kelapa sawit bulan Desember 2017 menggunakan *RapidMiner* pada tampilan *text view* hasil *cluster* sebagai berikut :



Gambar 10. Tampilan *Cluster Model* Data Pengujian
 Dalam bentuk *text view* dapat dilihat secara global bahwa anggota dari *cluster 0* berjumlah sebanyak 12 *items* yang terdiri dari nomor data 1,4,5,6,7,8,9,13,14,16,18,20 dengan keterangan baik , *cluster 1* berjumlah 13 *items* yang terdiri dari nomor data 11,12,15,19,21,22,23,24,26,27,28,29,30 dengan keterangan sangat baik dan *cluster 2* berjumlah 6 *items* yang terdiri dari nomor data 2,3,10,17,25,31 dengan keterangan kurang baik, total number *items* sebanyak 31.

IV. Kesimpulan

Metode *K-Means Clustering* berhasil diterapkan untuk mengelompokkan kualitas minyak kelapa sawit yang terdiri dari kriteria baik, sangat baik dan kurang baik dengan menggunakan *software RapidMiner*. Mendapatkan hasil pengelompokkan bahwa *cluster 0* (C_0) untuk kualitas baik terdapat 12 data, *cluster 1* (C_1) untuk kualitas sangat baik terdapat 13 data dan *cluster 2* (C_2) untuk kualitas kurang baik terdapat 6 data.

Daftar Pustaka

- [1] Krisdiarto, A.W., Sutiarto, Lilik., Widodo, K.H. (2017). “*Optimasi Kualitas Tandan Buah Segar Kelapa Sawit Dalam Proses Panen-Angkut Menggunakan Model Dinamis*”, *AGRITECH*, Vol. 37, No. 1, Februari 2017, ISSN : 0216-0455. <https://doi.org/10.22146/agritech.17015>
- [2] Muzakir, Ari.,Wulandari, R.A. (2016). “*Model Data Mining Sebagai Prediksi Penyakit Hipertensi Kehamilan Dengan Teknik Decision Tree*”, *Scientific Journal of Informatics* Vol. 3, No. 1, Mei 2016 ISSN : 2407-7658. <https://doi.org/10.15294/sji.v3i1.4610>
- [3] Sulastri, Heni., Gufroni, A.I. (2017). “*Penerapan Data Mining Dalam Pengelompokan Penderita Thalassemia*”, *Jurnal Teknologi dan Sistem Informasi - Vol. 03 No. 02 (2017)* 299-305. <https://doi.org/10.25077/TEKNOSI.v3i2.2017.299-305>
- [4] Nasari, Fina., Sianturi, C.J.M. (2016). “*Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat*”, *Cogito Smart Journal/Vol. 2/No. 2/Desember 2016*. [https://doi.org/2016:2\(2\):108-119](https://doi.org/2016:2(2):108-119)
- [5] Nishom.M.,Fathoni,M.Y.(2018).“*Implementasi Pendekatan Rule-Of-Thumb untuk Optimasi Algoritma K-Means Clustering*”, *Jurnal Pengembangan IT* Vol. 3, No.2. <https://doi.org/10.30591/jpit.v3i2.909>
- [6] Sulastri, Heni., Gufroni, A.I. (2017). “*Penerapan Data Mining Dalam Pengelompokan Penderita I. THALASSEMIA*”, *JURNAL TEKNOLOGI DAN SISTEM INFORMASI - VOL. 03 NO. 02 (2017)* 299-305. [HTTPS://DOI.ORG/10.25077/TEKNOSI.v3i2.2017.299-305](https://doi.org/10.25077/TEKNOSI.v3i2.2017.299-305)