Scientia Sacra: Jurnal Sains, Teknologi dan Masyarakat
Vol. 2 No. 3, September 2022

Contents lists available at Scientia Sacra

Scientia Sacra: Jurnal Sains, Teknologi dan Masyarakat
journal homepage: http://pijarpemikiran.com/index.php/Scientia

Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Siswa Baru pada SMK Letris Indonesia 1 Berbasis Web

Nurliana Anggraini[1], Lisda Fitriana Masitoh[2]
[1,2]Teknik Informatika, Universitas Pamulang
[1]nurliananggraini@gmail.com, [2]dosen01928@unpam.ac.id

Kata kunci:
Clustering, *K-means*, SMKLetris Indonesia 1, *web.*

Abstrak

Perkembangan teknologi saat ini seperti badai kebutuhan akan informasi semakin meningkat dan pengolahan data dalam berbagai aspek kehidupan manusia menjadi sangat penting. Perkembangan teknologi juga sangat berpengaruh pada sektor pendidikan. Melalui wawancara ditemukan fakta bahwa panitia PPDB SMK Letris Indonesia 1 masih melakukan pendataan manual dengan mengolah data satu persatu. Tujuan pengelompokkan adalah untuk menerapkan algoritma k-*means* dalam mengelompokkan kelas dari suatu objek sehingga data menjadi lebih terstruktur. metode yang digunakan pada penelitian ini menggunakan metode *clustering* dengan algoritma k-*means.* Hasil dari penelitian ini adalah mampu membantu SMK Letris Indonesia 1 dalam mengelola data menjadi lebih terstruktur.

Pendahuluan

Perkembangan teknologi saat ini seperti badai kebutuhan akan informasi semakin meningkat dan pengolahan data dalam berbagai aspek kehidupan manusia menjadi sangat penting[1]. Teknologi informasi adalah teknologi yang dibangun dengan teknologi komputer. Perkembangan yang terus berlanjut dari teknologi membawa aplikasi utama teknologi ini pada proses pengolahan data yang berujung pada informasi. Penyebab teknologi berkembang pesat hingga saat ini adalah kebutuhan manusia yang terus meningkat. Daya dorong dibalik kemajuan teknologi informasi dan komunikasi yang tidak dapat dipisahkan dari kehidupan sehari-hari. Perkembangan teknologi juga sangat berpengaruh pada sektor pendidikan[2]. Dengan siswa yang menjadi komponennya.

Dengan latar belakang masalah yang ada dapat dirumuskan beberapa permasalahan, yaitu: banyaknya data yang belum tersistem menyulitkan untuk mengolah data dalam satu waktu. Rumusan masalah yang ditemukan dalam

deskripsi judul adalah bagaimana penerapan metode *clustering* dengan algoritma *k-means* pada pengelompokan data calon siswa SMK Letris Indonesia 1?. Berdasarkan permasalahan yang diuraikan pada identifikasi masalah, maka penulis membatasi permasalahan untuk pengelompokan data menggunakan aplikasi *rapid miner*. Tujuan pengelompokkan data adalah untuk memperkirakan dan mengelompokkan kelas dari suatu objek sehingga data menjadi lebih terstruktur. Karena belum adanya pengelompokkan data yang dapat mempermudah dalam mengolah data yang dibutuhkan. Dengan adanya masalah dalam pengelompokkan data. Maka pengelompokkan data akan menggunakan algoritma *k- means clustering*.

Berdasarkan uraian latar belakang yang ada peneliti bermaksud untuk menerapkan metode *clustering* dengan algoritma *k-means* pada pengelompokkan data calon siswa baru pada SMK Letris Indonesia 1 berbasis web. Hal ini dikarenakan pentingnya suatu aplikasi untuk membantu proses pengelompokkan data siswa. Berdasarkan wawancara yang dilakukan peneliti di sekolah. Diketahui bahwa sekolah tersebut belum mempunyai aplikasi. Aplikasi berbasis web dengan tujuan untuk memudahkan panitia PPDB dalam mengelompokkan data siswa. Diharapkan dengan aplikasi ini SMK Letris Indonesia 1 dapat lebih mudah dalam mengelompokkan data siswa dan dapat mempersingkat waktu dalam mengolah data yang ada.

Metode

Pada penelitian ini peneliti menggunakan metode *clustering*. Metode *clustering* dengan algoritma k-*means*. Yang dapat digunakan untuk mengelompokkan sejumlah objek sesuai dengan atribut yang sama atau karakteristik yang sama ke dalam sejumlah kelompok. K-*means clustering* adalah metode analisis data atau metode data *mining* yang melakukan proses pemodelan tanpa pengawasan (*unsupervised*) dan merupakan metode yang melakukan pengelompokan data dengan sistem partisi[3]. Sistem ini mempermudah dalam proses pengelompokkan data calon siswa baru SMK Letris Indonesia 1 dengan menguraikan kegiatan yang dihasilkan pada tahapan perancangan. Adapun analisa kebutuhan yang dilakukan dengan melakukan wawancara dan pencatatan di SMK Letris Indonesia

Hasil dan Pembahasan

Pengumpulan data pada proses pengumpulan data ada 3 parameter yang digunakan dalam pengolahan data yaitu nilai tes tertulis, nilai wawancara dan nilai ujianakhir sekolah (UAS). Dengan data yang berjumlah 75 *record*.

Pengukuran jarak dengan metode *euclidean* dapat dituliskan dengan persamaan berikut:

$$j(v_1, v_2) = \sqrt{\sum_{k=1}^{N} \left(v_1(k) - v_2(k)\right)^2}$$

dimana v1 dan v2 adalah dua vektor yang jaraknya akan dihitung dan N menyatakan panjang vektor[4].

Adapun langkah dalam *cluster* dengan algoritma *K-Means* yaitu :

1. Menentukan jumlah *cluster* menentukan jumlah *cluster* yang digunakan pada data nilai tes tertulis, wawancara, dan UAS SMK Letris Indonesia 1. Sebanyak 3 *cluster* diantaranya Lulus, Lulus Berprestasi dan Tidak Lulus.
2. Menentukan *centroid* penentuan pusat awal *cluster (centroid)* ditentukan secara

random atau acak yang diambil dari data yang ada.

3. Menghitung jarak dari *centroid* menghitung jarak antara titik centroid dengan titik tiap objek dengan menggunakan *Euclidian Distance*. Adapun penghitungan *centroid* awal secara manual. Perhitungannya adalah sebagai berikut :

$$\boldsymbol{D_{(if)}} = \sqrt{(\boldsymbol{X_{il}} - \boldsymbol{X_{lj}})^2 + (\boldsymbol{X_{2i}} - \boldsymbol{X_{2j}})^2} =$$

Tabel 1. Menentukan *centroid*

Cluster	Tes tertulis	Wawancara	UAS
1	90	86	85
2	64	69	66
3	46	63	61

Tabel 2. Perhitungan jarak terpendek

No.	Nama	Tes tertulis	wawancara	UAS	C1	C2	C3	JARAK TERPENDEK
1	Akhyar Maulana Nasution	89	90	75	11	34	53	11
2	Aditya Dwi Indra	90	71	95	18	39	22	18
3	Aldo Rafi Alfian	70	75	80	23	16	12	12
4	Alvin Pradita Kwan	45	65	59	56	21	39	21
5	Alvino Dwi Putra	65	75	53	42	14	27	14
6	Alya Putri Shifani	80	70	75	21	18	0	0
7	Andhika Barru Hijriyan	90	85	81	4	34	19	4
8	Andrie Fahrurzy	70	70	73	28	9	10	9
9	Anjani Dwiyana	96	93	85	9	44	30	9
10	Azzelia Lubna R	60	55	48	57	23	37	23
11	Berryl Reivan Arkananta	45	60	58	59	22	40	22
12	Bima Aji Nugroho	60	70	72	36	7	20	7
13	Desya Christy	85	90	88	7	37	24	7
14	Dimas Allifahmi	52	68	55	52	16	34	16

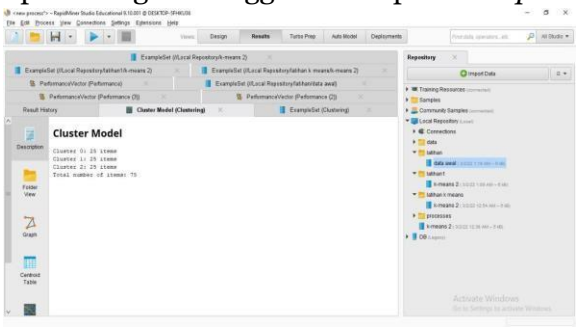
15	Dinda Indah Rahmah	40	60	70	58	26	42	26
16	Dinda Nur Ardila	89	90	75	11	34	22	11
17	Diyas Salam Arumi	90	71	95	18	39	22	18
18	Egan Nabhan Tri Aldianka	70	75	80	23	16	12	12
19	Eva Lailatu Syarifah	45	65	59	56	21	39	21
20	Evangelia Zafirah Ramadhani	65	75	53	42	14	27	14
21	Fahri Nur Alim	80	70	75	21	18	0	0
22	Faiz Sulaiman	90	85	81	4	34	19	4
23	Fajar Muhammad Aryansyah	70	70	73	28	9	10	9
24	Fajra Akbar Ilhamsyah	96	93	85	9	44	30	9
25	Farel Wahyu Wijanarko	60	55	48	57	23	37	23
26	Fariziah Alikha Noveliza	45	60	58	59	22	40	22
27	Fauza Hedistira Firly	60	70	72	36	7	20	7
28	Febrian Saputra	85	90	88	7	37	24	7
29	Fiola Cahyani Regita	52	68	55	52	16	34	16
30	Gerald Adriel Palimbong	40	60	70	58	26	42	26
31	Harsha Dimas Syahputra	89	90	75	11	34	22	11
32	Hasyim	90	71	95	18	39	22	18
33	Hilmi Fata Fadhila	70	75	80	23	16	12	12
34	Iqbal Kurniawan	45	65	59	56	21	39	21
35	Irfan Aziz Anggara	65	75	53	42	14	27	14
36	Isfal Abidal Aziz	80	70	75	21	18	0	0
37	Kayla Luvena Kazimir	90	85	81	4	34	19	4
38	Kayra Vanya Nanditha	70	70	73	28	9	10	9
39	Krisna	96	93	85	9	44	30	9

	Angga Pratama							
40	Kukuh Wicaksono	60	55	48	57	23	37	23
41	Lingga Arya Saputra	45	60	58	59	22	40	22
42	Lukman Nul Hakim	60	70	72	36	7	20	7
43	Lukman Nurhakim	85	90	88	7	37	24	7
44	M Afreza Attar	52	68	55	52	16	34	16
45	Muhamad Andi Sallahudin	40	60	70	58	26	42	26
46	Muhammad Akbar Fauzan	89	90	75	11	34	22	11
47	Muhammad Al Zikri	90	71	95	18	39	22	18
48	Muhammad Arya Naratama	70	75	80	23	16	12	12
49	Muhammad Fatur Rahman	45	65	59	56	21	39	21
50	Muhammad Gathan A	65	75	53	42	14	27	14
51	Muhammad Iskandar Dinata	80	70	75	21	18	0	0
52	Muhammad Ragil Java. Rangk	90	85	81	4	34	19	4
53	Muhammad Refi Andika	70	70	73	28	9	10	9
54	Muhammad Reza Fitrahul Ak	96	93	85	9	44	30	9
55	Muhammad Ridwan	60	55	48	57	23	37	23
56	Nabilah	45	60	58	59	22	40	22
57	Nayla Putri Asnarani	60	70	72	36	7	20	7
58	Neo Cahaya Budiarto	85	90	88	7	37	24	7
59	Nopal Ardiansyah	52	68	55	52	16	34	16
60	Pandega Zufar Ayusman Arya	40	60	70	58	26	42	26
61	Paskalis Briniant Pramono	89	90	75	11	34	22	11
62	Rama Gumilang	90	71	95	18	39	22	18

63	Razi Shafvello	70	75	80	23	16	12	12
64	Regina Serennita	45	65	59	56	21	39	21
65	Reyhan Wafa Adzurri	65	75	53	42	14	27	14
66	Rifky Setiyawan	80	70	75	21	18	0	0
67	Rifqi Pramadi	90	85	81	4	34	19	4
68	Suhal Pani	70	70	73	28	9	10	9
69	Syafiq Nur Hidayah	96	93	85	9	44	30	9
70	Teguh Nur Hidayat	60	55	48	57	23	37	23
71	Valtteri Sebastian Vanhatalo	45	60	58	59	22	40	22
72	Vanessa Noela Wirawan	60	70	72	36	7	20	7
73	Vistya Dian Lestari	85	90	88	7	37	24	7
74	Zhauqeeza Axcel Putra	52	68	55	52	16	34	16
75	Zidan Achilla Muhammad	40	60	70	58	26	42	26

Karena pada Iterasi ke-2 dan ke-3 posisi *cluster* tidak berubah maka iterasi dihentikan dan hasil akhir yang diperoleh yaitu ; C1 memiliki 25 anggota, C2 memiliki 40 anggotadan C3 memiliki 10 anggota.

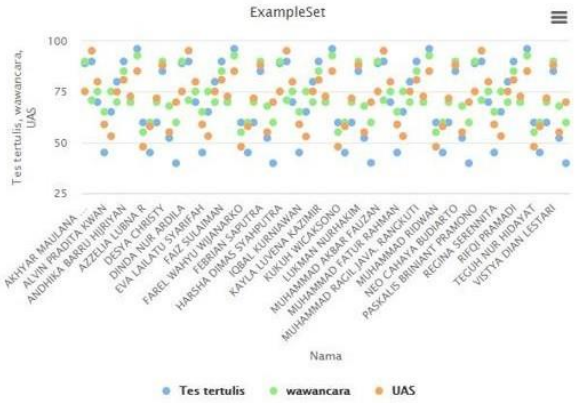
Setelah melakukan perhitungan manual maka hasil yang ada akan dibandingkandengan perhitungan menggunakan aplikasi *rapid miner*.



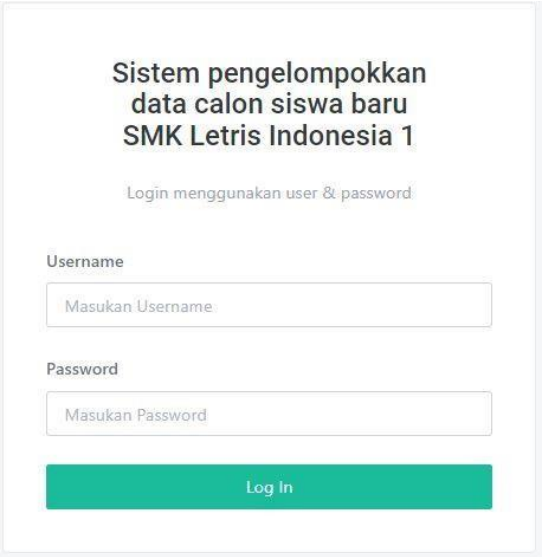
Gambar 1. Hasil *cluster* dengan aplikasi *rapid miner*.

Melalui aplikasi *rapid miner* anggota *cluster* yang ada berbeda dengan hasil yang dihitung manual. Karena memiliki tingkat akurasi yang lebih tinggi. Dengan masing-masing anggota pada *cluster*: *cluster* 0 memiliki 25 anggota, *cluster* 1 memiliki 25 anggota dan *cluster* 3 memiliki 25 anggota.

Gambar 2. Hasil dalam bentuk *scatter/buble*.

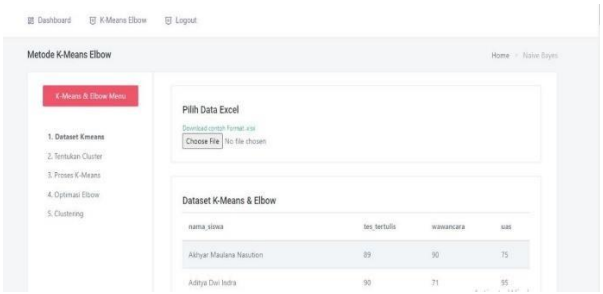


Setelah melakukan perbandingan perhitungan manual dengan aplikasi *rapid miner*. Selanjutnya akan dikembangkan menjadi *website*. Berikut ini implementasi dari *form login,* dimana *form* ini digunakan oleh admin untuk masuk ke dalam sistem pengelompokkan data calon siswa baru SMK Letris Indonesia 1.



Gambar 3. Halaman *login* admin PPDB

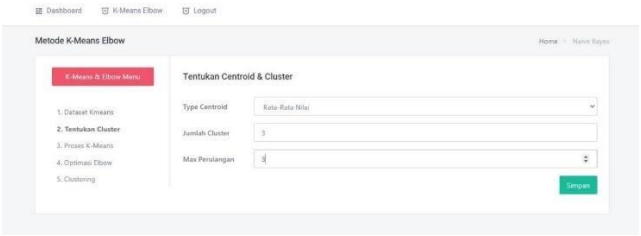
Berikut ini adalah menu utama dari sistem yang akan ditampilkan apabila sudah melakukan *login*. pada menu ini menu bar *Dataset Kmeans*, Tentukan *Cluster*, Proses *Kmeans, Optimasi Elbow* dan *Clustering*. Fungsi *form Dataset kmeans* adalah untuk memilih data yang akan dihitung. Dengan tipe file excel extensi .xlxs



Gambar 4. Halaman awal

Form ini adalah *form* yang muncul jika admin melakukan klik pada menu Tentukan *Cluster*. Terdapat beberapa *field yaitu:*

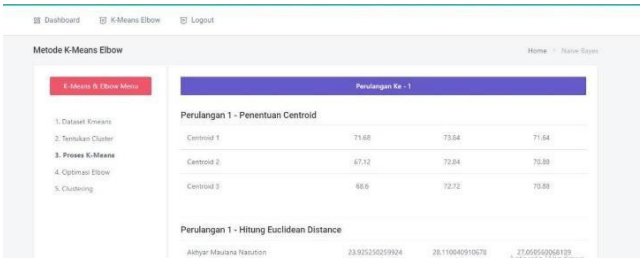
1) *Type Centroid* untuk memilih antara rata-rata nilai, *random centroid, customcentroid* dan isi manual.
2) Jumlah *cluster* untuk menentukan jumlah pada setiap *cluster*
3) *Max* perulangan untuk menentukan maksimal perulangan pada setiap *cluster*.



Gambar 5. Tentukan *cluster*

Berikut ini adalah *form* Proses *K-means* yang akan muncul apabila admin melakukan

input pada *form* tentukan *cluster*. Hasil dari proses akan ditampilkan pada menu ini.



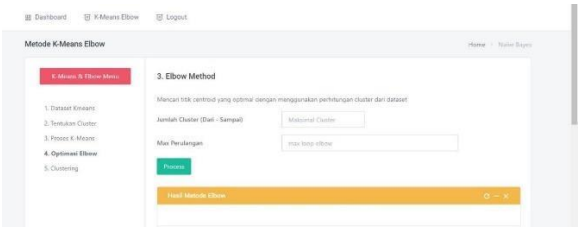
Gambar 5. Proses *kmeans*

Form ini akan muncul jika admin melakukan klik pada *optimasi elbow*. Dengan *field*

sebagai berikut:

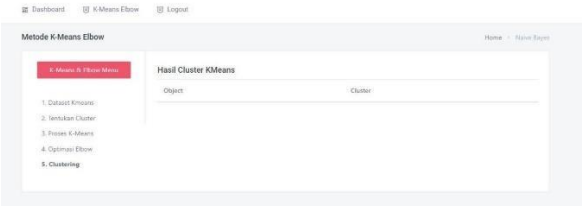
1) Jumlah *cluster* untuk menentukan jumlah dari *cluster* yang akan dihitung.
2) *Max* perulangan untuk menentukan maksimal perulangan pada *cluster* yang

akan dihitung.



Gambar 6. Optimasi *elbow*

Form ini akan tampil jika admin melakukan klik pada menu *Clustering*. Dalam menu *clustering* akan ditampilkan hasil cluster kmeans dengan tampilan *object* dan *cluster*.



Gambar 7. C*lustering*

Kesimpulan

Menerapkan metode *clustering* untuk pengolahan data di SMK Letris Indonesia 1 dengan menggunakan metode *euclian distance* dengan cara tersebut dapat ditemukan kemudahan. Dalam menerapkan metode *euclian distance* untuk menghitung jarak terpendek dari masing-masing nilai menggunakan rumus metode *euclian distance*. Dengan menghitung jarak dua titik dengan mengetahui nilai masing-masing atribut pada kedua titik tersebut. Dalam hasil data analisa yang dihitung secara manual menggunakan rumus metode *euclian distance* mendapatkan hasil yang diperoleh yaitu: C1 memiliki 25 anggota yang diartikan bahwa kelompok pertama adalah kategori kemampuan peserta didik lulus berprestasi, C2 memiliki 40 anggota yang diartikan bahwa kelompok kedua adalah kategori kemampuan peserta didik lulus dan C3 memiliki 10 anggota yang diartikan bahwa kelompok ketiga adalah kategori kemampuan peserta didik tidak lulus.

Sedangkan dalam perhitungan yang dilakukan melalui aplikasi dengan tingkat akurasi lebih tinggi. Dan melakukan pengujian pada setiap kelompok yang berbeda dapat ditentukan hasil yang lebih baik sebagai berikut: C1 memiliki 25 anggota yang diartikan bahwa kelompok pertama adalah kategori kemampuan peserta didik lulus berprestasi, C2 memiliki 25 anggota yang diartikan bahwa kelompok kedua adalah kategori kemampuan peserta didik lulus dan C3 memiliki 25 anggota yang diartikan bahwa kelompok ketiga adalah kategori kemampuan peserta didik tidak lulus.

Daftar Pustaka

[1] Asroni, A., Fitri, H., & Prasetyo, E. "Penerapan Metode *Clustering* dengan Algoritma K-*Means* pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)". Semesta Teknika, 21(1)., Vol. 21, No.1 pp. 60-64, 2018.

[2] Johan Oscar Ong. "Implementasi Algoritma K-*Means Clustering* Untuk Menentukan Strategi *Marketing* President University". Jurnal Ilmiah Teknik Industri, 12(1), Vol. 12, No. 1 pp 10–20, 2022.

[3] Gustientiedina, G., Adiya, M. H., & Desnelita, Y. "Penerapan Algoritma K-*Means* Untuk *Clustering* Data Obat-Obatan". Jurnal Nasional Teknologi Dan Sistem Informasi, 5(1), Vol. 5, No. 1 pp 17–24, 2019.