

APPLICATION OF THE K-MEANS METHOD FOR PREPARATION OF NEW STUDENTS ENTRY SELECTION

PENERAPAN METODE K-MEANS UNTUK PERSIAPAN SELEKSI MASUK MAHASISWA BARU

Nur Fuad¹, Listiarini Edy Sudiati², Ninik Haryani³

¹Stmik Aki Pati
Jln. Kamandowo 13 Pati
Kidul, Pati, Jawa Tengah
nurfuad640@gmail.com

²Stmik Aki Pati
Jln. Kamandowo 13 Pati
Kidul, Pati, Jawa Tengah
listiarini@gmail.com

³Stmik Aki Pati
Jln. Kamandowo 13 Pati
Kidul, Pati, Jawa Tengah
ninik.winong@gmail.com

ABSTRACT

The moment of acceptance of new students is often an event for schools to get potential students, especially in the academic field. This is done because the good name of the school is strongly influenced by the achievements of its students either through competitions, art events or other activities. But on the other hand, many prospective new students do not understand what potential is in them. For that, we need a system that is able to classify students based on student potential. The grouping uses the National Examination scores as the basis for this grouping. The clustering method used is the K-Means Clustering method. The data is divided into two clusters which are divided into general school groups and madrasah school groups. From the results of clustering, it was found that 6 students were in the general school group and 11 students in the madrasa school group.

Keywords: new students, clustering, K-means

1. LATAR BELAKANG

Penggalian potensi peserta didik sangatlah penting terutama untuk masa depan masing-masing peserta didik itu sendiri. Melalui penggalian potensi ini, guru atau pihak sekolah bahkan wali siswa dapat mengetahui potensi atau bakat yang dimiliki oleh siswa.^[1] Pengetahuan ini sering kali terlambat atau bahkan tidak terdeteksi sama sekali sehingga siswa sulit berkembang karena kebanyakan siswa akan merasa “dipaksa” untuk melakukan atau mempelajari hal yang tidak disukai dan berujung pada kenakalan seperti pergi *membolos* saat jam pelajaran.^{[1][2]}

Manajemen potensi dilakukan mengacu pada memikat, mengembangkan, memotivasi, dan mempertahankan bakat siswa secara efisien dalam lingkup akademik maupun non-akademik. Pengembangan potensi bertujuan untuk melatih siswa agar memiliki seperangkat pengetahuan, keterampilan, dan kemampuan yang kompleks seperti keterampilan membuat keputusan, kepercayaan diri, pemikir kreatif, *knowledge-seeking*, *relationship-building*, *public-speaking*, serta manajemen resiko.^[3] Tentunya, bakat atau potensi harus diselaraskan dengan minat dari siswa itu sendiri. Sehingga dengan adanya studi ini siswa, orang tua, guru, maupun pihak sekolah dapat menyusun strategi lebih lanjut guna mengembangkan bakat atau potensi siswa secara optimal tanpa menghalangi minat siswa.

Salah satu sistem yang dapat dikembangkan yaitu dengan algoritma K-Means. Dimana dataset calon siswa dapat dikelompokkan secara otomatis sesuai dengan kriteria yang sudah ditentukan. Pada penelitian sebelumnya yang membahas tentang aplikasi algoritma K-means memiliki kelebihan dan kekurangan masing-masing, dimana clustering K-means digunakan untuk analisa hasil ujian mahasiswa.^[4] Dalam penelitian tersebut, algoritma clustering k-means dipadukan dengan metode *elbow* untuk menentukan jumlah *clustering*, sehingga kinerja sistem lebih efisien, namun menyebabkan *noise* yang cukup besar karena tidak dilakukan normalisasi data. Normalisasi data selain untuk mengurangi *noise* juga berguna untuk menekan kemungkinan adanya kesalahan data.^[5]

Selain melakukan normalisasi data, penggunaan metode pengelompokan juga harus disesuaikan dengan kebutuhan, karena metode yang digunakan akan berpengaruh pada kinerja sistem. Efisiensi

juga menjadi menjadi faktor penting dalam pemilihan penggunaan metode karena berhubungan dengan kemampuan perangkat dalam mengolah sejumlah data. Misal, penggunaan pengelompokan K-means dengan metode *Euclidean Distance*, menghasilkan kinerja sistem cukup baik, dimana sistem mampu untuk mengolah data yang cukup lebar yaitu 3 variabel dengan jumlah responden 8.741 mahasiswa.^[6]

Dari pembahasan diatas maka perlu dikembangkan suatu sistem pendukung yang mampu mengelompokkan siswa siswa berdasarkan potensi akademiknya menggunakan algoritma K-Means dengan metode *Euclidean Distance*.^[7] Variabel yang digunakan adalah nilai Ujian Nasional pada jenjang sebelumnya.

2. METODOLOGI

a. K-Means Clustering

Algoritma *K-Means clustering* adalah algoritma pengelompokan tradisional yang termasuk kedalam *unsupervised machine learning*, dirancang oleh Macqueen sebagai metode pengelompokan sederhana dan efisien. Meski sederhana, metode ini memiliki keunggulan skalabilitas dan efisiensi tinggi untuk memproses kumpulan data besar. Algoritma pengelompokan K-Means memiliki cakupan aplikasi yang luas.^{[5][8]} Penerapan metode K-Means dapat digunakan dalam banyak bidang seperti aplikasi biologi, analisa citra, riset pasar maupun bidang lainnya. Metode *K-means* juga dapat diintegrasikan dengan banyak platform terbaru seperti *Internet of Thing*.^[8]

K-means clustering efektif digunakan untuk mempartisi kumpulan data yang diberikan ke dalam k grup atau k cluster, di mana k mewakili jumlah grup atau cluster. Jumlah cluster dihitung oleh analis untuk mendapatkan pengelompokan data yang efektif.^[9] Metode *k-means clustering* memisahkan data kedalam cluster sedemikian rupa sehingga nilai data dalam *cluster* yang sama mirip satu sama lain. Dalam pengelompokan k-means, setiap *cluster* diwakili oleh titik pusat atau meannya yang juga disebut sebagai *centroid* dan dihitung sebagai nilai rata-rata dari data dalam cluster tersebut.^[10]

b. Methodology of K-Means Clustering

Langkah-langkah melakukan clustering dengan *K-Means* adalah:^{[10][11]}

- 1) Menentukan jumlah cluster yang ingin dibentuk, dimana nilai K adalah banyaknya cluster;
- 2) Menentukan *cluster centroid* awal. Centroid awal tergantung secara acak dari data, dan jumlah centroid sama dengan jumlah cluster;
- 3) Setelah *centroid* awal ditentukan, maka setiap data akan dihitung jaraknya dengan *centroid* terdekat menggunakan persamaan korelasi antara dua objek *Euclidean Distance*;

4) [1]

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

- 5) Setelah menghitung jarak data dengan *centroid*, langkah berikutnya adalah mengelompokkan data berdasarkan jarak minimum. Suatu data akan menjadi anggota cluster yang terdekat (kecil) dari cluster tersebut;
- 6) Berdasarkan *clustering* awal, selanjutnya mencari *centroid* baru berdasarkan keanggotaan dari masing-masing cluster dengan menghitung rata-rata cluster;^[2]

$$v = \frac{\sum_{i=1}^n x_i}{n} \quad ; i = 1, 2, 3, \dots, n.$$

Meski memiliki keuntungan dalam pengolahan *big-data*, metode K-means memiliki kekurangan dimana tidak dapat diketahui *real cluster*, karena apabila dimasukan dengan cara yang berbeda, berpeluang muncul *cluster* yang berbeda. Kontribusi dari atribut juga tidak dapat diketahui karena dianggap bahwa setiap atribut memiliki bobot yang sama.

c. Case Study and Data Collecting

Pada tahun 2021, Madrasah Ibtidaiyah Tarbiyatus Salam memiliki 17 peserta didik yang telah melaksanakan Ujian Nasional dan semua dinyatakan lulus. Namun, setelah kelulusan mereka, banyak dari siswa tersebut yang masih bimbang untuk menentukan apakah akan melanjutkan ke sekolah umum atau tetap disekolah yang berbasis madrasah. Untuk itu, perlu dibuat sebuah sistem pendukung yang mampu memberikan rekomendasi untuk para siswa melanjutkan sekolah. Dalam sistem ini, rekomendasi didasarkan pada nilai ujian siswa yang terdiri dari mata pelajaran Al Qur'an-Hadits, Aqidah-Akhlaq, Fiqh, dan Bahasa Arab sebagai variabel sekolah agama. Sedangkan untuk variabel sekolah umum diambil dari nilai ujian mata pelajaran Bahasa Indonesia, Matematika, IPA, dan IPS.

3. HASIL DAN PEMBAHASAN

Proses identifikasi masalah menjadi sangat penting pada tahap analisis kebutuhan sistem. Dari hasil identifikasi didapatkan kriteria –kriteria yang harus bisa dipenuhi oleh sistem sehingga sistem akan secara tepat dalam memecahkan atau membantu menyelesaikan masalah pengguna. Penggunaan aplikasi K-means didasari karena kemudahan dan efisiensinya dalam mengolah banyak data. Hasil analisa kebutuhan menunjukkan kebutuhan pada sistem yang dapat mengklasifikasikan siswa pada kelompok ilmu pengetahuan umum dan ilmu pengetahuan agama. Dimana pada kelompok (*cluster*) ilmu pengetahuan umum akan direkomendasikan untuk melanjutkan sekolah di sekolah umum. Sedangkan kelompok ilmu pengetahuan agama direkomendasikan untuk melanjutkan disekolah berbasis Madrasah. Sistem juga dirancang agar pengguna dapat melakukan pengelompokkan siswa dengan mudah dan sederhana. Dengan adanya pengelompokkan ini diharapkan mampu meningkatkan prestasi siswa dan sekolah.



Sistem Pengelompokan Siswa Lulusan MI Tarbiyatus Salam Tahun 2021
Penerapan Metode K-Means

Jumlah Kolom Penilaian
8

Jumlah Siswa
17

Proses

Gambar 1. Halaman Input Data Siswa

Untuk mengantisipasi kesalahan pada saat proses *entry-data* maka disediakan fitur edit data. Data yang ada direncanakan untuk dibagi menjadi dua *cluster* dimana *cluster* 1 merupakan kelompok siswa yang nantinya direkomendasikan untuk melanjutkan ke sekolah umum, sedangkan siswa yang tergabung di *cluster* 2 akan direkomendasikan untuk melanjutkan ke sekolah berbasis madrasah.

Sistem Pengelompokan Siswa Lulusan MI Tarbiyatus Salam Tahun 2021									
Penerapan Metode K-Means									
#	Qur'an-Hadits	Aqidah Akhlaq	Fiqh	B. Arab	B. Indonesia	Matematika	IPA	IPS	centroid
Adi	65	69	70	65	69	60	077	69	<input type="checkbox"/>
Ahmad	82	85	92	85	84	65	74	73	<input checked="" type="checkbox"/>
Dwiva	95	87	97	93	95	85	91	83	<input type="checkbox"/>
Feronika	71	79	74	76	80	68	81	65	<input type="checkbox"/>
Fitri	73	70	73	65	69	60	65	65	<input type="checkbox"/>
Fransisca	70	80	87	73	83	67	81	65	<input type="checkbox"/>
Jangkung	69	66	65	65	65	60	66	65	<input type="checkbox"/>
Keyla	68	80	67	65	69	65	65	65	<input type="checkbox"/>
Meisya	93	89	87	78	80	67	75	65	<input type="checkbox"/>
Muhyi	75	82	67	69	65	60	78	65	<input type="checkbox"/>
Nabila	89	83	89	74	77	64	75	65	<input type="checkbox"/>
Rizky	92	80	89	75	85	78	82	75	<input type="checkbox"/>
Saskia	68	72	74	65	65	60	76	65	<input type="checkbox"/>
Sesilia	72	82	79	65	65	61	67	67	<input type="checkbox"/>
Siti K.	81	80	84	71	79	70	76	65	<input checked="" type="checkbox"/>
Siti W.	76	91	72	65	72	61	77	65	<input type="checkbox"/>
Vita	90	91	89	87	82	71	82	79	<input type="checkbox"/>

Gambar 2. Data Siswa

Dengan data tersebut, diambil secara acak untuk mewakili masing-masing *cluster* data untuk dijadikan sebagai *centroid* awal. Pembagian *cluster* diharapkan mampu membagi data menjadi kelompok akademik yang memiliki nilai akademik sangat baik dan kelompok non-akademik dengan nilai yang cukup. Sehingga data yang digunakan yaitu data dengan nama “Ahmad” untuk *cluster* 1 dan “Siti K.” untuk *cluster* 2. Sehingga nilai untuk *centroid* awal adalah:

$$C1 = (82, 85, 92, 85, 84, 65, 74, 73)$$

$$C2 = (81, 80, 84, 71, 79, 70, 76, 65)$$

Setelah *centroid* awal ditentukan maka dapat ditentukan jarak antara setiap data yang ada dengan terhadap semua *centroid* menggunakan persamaan *Euclidean Distance*. Dari perhitunga didapati;

Tabel 1. Jarak Data Lulusan MI Tarbiyatus Salam Tahun 2021 dengan Masing-Masing *Centroid*

Data	C1	C2
Adi	41.20	28.59
Ahmad	0.00	19.99
Dwiva	34.00	44.15
Feronika	26.01	15.68
Fitri	38.23	25.12
Fransisca	21.62	13.97
Jangkung	46.23	33.13
Keyla	40.52	27.38
Meisya	16.82	17.42

Muhyi	36.90	24.75
Nabila	17.01	12.17
Rizky	21.29	19.95
Saskia	39.37	25.86
Sesilia	33.68	22.29
Siti K	19.99	0.00
Siti W	33.14	21.38
Vita	15.74	27.13

Setelah didapati jarak data dengan masing-masing *centroid* selanjutnya data dapat di kelompokkan kedalam *cluster* yang memiliki jarak terdekat dengan data. Penentuan pengelompokan data dihitung dari jarak data dengan rata-rata data pada *cluster* yang sama. Pada iterasi pertama, *cluster* 1 memiliki nilai rata-rata data 28,34 sedangkan nilai rata-rata data pada *cluster* 2 adalah 22,29. *Cluster* ini merupakan *cluster* sementara karena proses pengelompokan masih menggunakan *centroid* awal yang dipilih secara acak. Proses pengelompokan masih harus melalui beberapa kali iterasi agar hasil akhir dapat mencapai iterasi maksimum, yang artinya pengelompokan mencapai hasil dimana kelompok data sudah memiliki ciri khas masing-masing. Iterasi akan mencapai maksimum ketika tidak ada lagi data yang bergerak atau berubah posisi *cluster*-nya.

Table 2. *Clustering* Sementara Lulusan MI Tarbiyatus Salam Tahun 2021

Data	C1	C2	C1	C2
Adi	41.20	28.59	0	0
Ahmad	0.00	19.99	1	1
Dwiva	34.00	44.15	0	0
Feronika	26.01	15.68	1	1
Fitri	38.23	25.12	0	0
Fransisca	21.62	13.97	1	1
Jangkung	46.23	33.13	0	0
Keyla	40.52	27.38	0	0
Meisya	16.82	17.42	1	1
Muhyi	36.90	24.75	0	0
Nabila	17.01	12.17	1	1
Rizky	21.29	19.95	1	1
Saskia	39.37	25.86	0	0
Sesilia	33.68	22.29	0	1
Siti K	19.99	0.00	1	1
Siti W	33.14	21.38	0	1
Vita	15.74	27.13	1	0

Dengan didapati *clustering* sementara maka langkah selanjutnya adalah mengulangi iterasi dengan *centroid* yang diambil berdasarkan rata-rata data dari *cluster* yang sama saat *clustering* pada iterasi sebelumnya. Pengulangan proses atau iterasi dilakukan hingga sudah tidak ada lagi data yang bergeser dari *cluster*-nya. Pada penelitian ini didapati hasil akhir *cluster* 1 diisi 6 orang sedangkan *cluster* 2 bersisihan 11 orang, seperti pada gambar berikut;

Sistem Pengelompokan Siswa Lulusan MI Tarbiyatus Salam Tahun 2021			
Penerapan Metode K-Means			
DETAIL DATA		CENTROID TERDEKAT	
Jumlah Cluster	2	#	
Jumlah Baris	17	Cluster 1	Cluster 2
Jumlah Kolom	8	Ahmad	Yes
Centroid 1	90.17, 85.83, 90.5, 82, 83.83, 71.67, 79.8:	Dwiva	Yes
Centroid 2	71.64, 77.36, 73.82, 67.64, 71, 62.91, 73.1:	Meisya	Yes
		Nabila	Yes
		Rizky	Yes
		Vita	Yes
		Adi	Yes
		Feronika	Yes
		Fitri	Yes
		Fransisca	Yes
		Jangkung	Yes
		Keyla	Yes
		Muhyi	Yes
		Saskia	Yes
		Sesilia	Yes
		Siti K.	Yes
		Siti W.	Yes

Gambar 3. Hasil Akhir Clustering Data

4. KESIMPULAN

1. Dengan aplikasi berbasis *K-Means Clustering* pihak sekolah bisa dengan tepat dalam menyalurkan alumennya sesuai dengan kebutuhan dan kemampuan para siswa.
2. Lulusan MI Tarbiyatus Salam tahun 2021 diarahkan untuk melanjutkan ke sekolah umum sebanyak 35%, dan 65% diarahkan untuk melanjutkan ke sekolah berbasis madrasah.

DAFTAR PUSTAKA

- [1] McGrath C H, dan Frearson, M. (2016). *Foro de Educación*. **14:21** 167-195. Title Admission Systems and Student Mobility: A Proposal for an EU-Wide Registry for University Admission
- [2] Battaglia O R, Paola B D, dan Fazio C.(2017). K-means Clustering to Study How Student Reasoning Lines Can Be Modified by a Learning Activity Based on Feynman's Unifying Approach. *EURASIA Journal of Mathematics Science and Technology Education*. **13:6** 2005-2038.
- [3] Veluchamy R, Bera S, Rajput A S, dan Krishnan A R.(2016). Students' Talent Management in Creating Student CEOs. *International Journal of Control Theory and Applications*. **9:37** 685-692.
- [4] Aggarwal D, dan Sharma D. (2019). Application of Clustering for Student Result Analysis. *International Journal of Recent Technology and Engineering*. **7:6C** 50-53.
- [5] Chang W, Ji X, Liu Y, Xiao Y, Chen B, Liu H, dan Zhou S. (2020). Analysis of University Students' Behavior Based on a Fusion K-Means Clustering Algorithm. *Applied Sciences*. **10** 6566.

- [6] Muttaqien H, Lutfi M, KH Musliadi, Muis A, dan Zainuddin H. (2019). Recommendation of Student Admission Priorities Using K-Means Clustering. *EAI Publisher: 1st ICOST Proceeding*. DOI 10.4108/eai.2-5-2019.2284614.
- [7] Sya'iyah K, Yuliansyah H, dan Arfiani I. (2019). Clustering Student Data Based On K-Means Algorithms. *International Journal of Scientific & Technology Research*. **8:08** 1014-1018.
- [8] Chunsheng L, Kejia Z, Tao L, dan Yanan H.(2020). Analysis of Postgraduate Admission Scores Based on K-means Algorithm. *IEEE Publisher: 2020 ICBDIE*, 402-406. doi: 10.1109/ICBDIE50010.2020.00101.
- [9] Kurniawan C, Setyosari P, Kamdi W, dan Ulfa S. (2018). Electrical Engineering Student Learning Preferences Modelled Using K-Means Clustering. *Global Journal of Engineering Education*. **20:2** 140-145.
- [10] Sari H L, Suranti D, Zulita L N.(2017). Implementation of K-Means Clustering Method for Electronic Learning Model. *IOP Conference Series: Journal of Physics: Conference Series*. **930** 012021.
- [11] Purba W, Tamba S, dan Saragih J.(2018). The Effect of Mining Data K-Means Clustering Toward Students Profile Model Drop Out Potential. *IOP Conference Series: Journal of Physics: Conference Series*. **1007** 012049.