

PENENTUAN SISWA BERPRESTASI MENGGUNAKAN METODE K-MEANS CLUSTERING DI SMP TAKHASSUS AL QUR'AN SADAMIYYAH

Nina Ismaya¹, Akhmad Khanif Zyen²

^{1,2,3)}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Nahdlatul Ulama Jl. Taman Siswa(Pekeng) Tahunan, Jepara, Jawa Tengah

e-mail: ¹161240000479@unisnu.ac.id, ²khanif.zyen@unisnu.ac.id

Informasi Artikel

Diterima: 05-01-2022

Direvisi: 28-01-2022

Disetujui: 02-02-2022

Abstrak

Prestasi belajar siswa merupakan prestasi yang diperoleh siswa dalam hal pengetahuan, keterampilan, dan sikap berkat pengalaman dan latihan yang telah dilalui oleh setiap individu. Penilaian dilakukan dengan memberikan nilai yang dibuat oleh guru. Namun mengingat banyaknya data siswa, tidaklah mudah untuk mencari siswa berprestasi, seperti yang terjadi di SMP Takhassus Al Qur'an Sadamiyaah Bangsri Jepara. Misalnya, mata pelajaran takhassus biasanya diikuti sertakan dalam lomba-lomba di booth tingkat kabupaten, nasional dan internasional. Dalam menentukan rekomendasi siswa berprestasi pada mata pelajaran tersebut, diperlukan suatu metode untuk menyelesaikan masalah secara cepat, tepat, dan akurat. Dalam penelitian ini diambil 414 data dari data siswa kelas VIII semester genap tahun ajaran 2015/2016-2019/2020 yang memiliki beberapa atribut yaitu nama, nilai tugas, nilai UTS, nilai UAS pada setiap mata pelajaran takhassus, daftar kehadiran, dan sikap menghargai. Data tersebut kemudian diolah pada alat RapidMiner menggunakan metode clustering K-Means, serta Davies-Bouldin Index untuk evaluasi dan validasi hasil. Hasil perhitungan Indeks Davies-Bouldin adalah 0,673 pada Microsoft Excel dan 0,041 untuk RapidMiner. Pengelompokan cukup baik karena mendekati 0 (nol-negatif ≥ 0).

Kata Kunci: Prestasi Siswa, Mata Pelajaran Takhassus, Algoritma K-Means, Clustering, Indeks Davies-Bouldin, RapidMiner.

Abstract

Student learning achievement is an achievement obtained by students in terms of knowledge, skills, and attitudes thanks to the experience and training that each individual has gone through. The assessment is carried out by giving a grade made by the teacher. However, considering the large number of student data, it is not easy to find outstanding students, as is the case at SMP Takhassus Al Qur'an Sadamiyaah Bangsri Jepara. For example, the takhassus subject is usually included in competitions at booth the district, national and international levels. in determining the recommendations of outstanding students in these subjects, a method is needed to solve problems quickly, precisely, and accurately. In this study, 414 data were taken from the data of grade VIII students in the even semester of the 2015/2016-2019/2020 academic year which had several attributes, namely name, assignment value, UTS score, UAS score in each takhassus subject, list attendance, and value attitude. The data is then processed on the RapidMiner tool using the K-Means clustering method, as well as the Davies-Bouldin Index for evaluation and validation of the results. The results of the calculation of the Davies-Bouldin Index are 0.673 in Microsoft Excel and 0.041 for RapidMiner. The clustering is fairly good because it is close to 0 (zero-negative ≥ 0).

Keywords: Student Achievement, Takhassus Subjects, K-Means Algorithm, Clustering, Davies-Bouldin Index, RapidMiner.



1. Pendahuluan

Tingginya tingkat keberhasilan siswa merupakan cerminan daripada kualitas dunia pendidikan. Belakangan ini, dunia pendidikan dituntut untuk memiliki daya saing yang tinggi dengan memanfaatkan kualitas serta kuantitas sumber daya manusia (SDM) yang ada, seperti halnya siswa berprestasi. Namun realitanya, seseorang mempunyai kemampuan dalam hal penguasaan dan pemahaman yang berbeda-beda. Hal tersebut dipicu oleh perbedaan pengalaman serta pengetahuan yang dimilikinya. Prestasi belajar siswa merupakan hasil yang dicapai atau diperoleh siswa baik ilmu pengetahuan, ketrampilan, dan sikap berkat pengalaman serta latihan yang telah dilalui masing-masing individu [1]. Secara umum prestasi siswa dapat dinilai dari mata pelajaran yang diambil, baik teori maupun praktik. Selain itu juga penilaian ekstrakurikuler, dan daftar hadir siswa. Mengingat banyaknya data siswa, maka untuk mengetahui siswa berprestasi bukanlah hal mudah yang dilakukan oleh guru maupun sekolah, sama halnya di SMP Takhassus Al Qur'an Sadamiyyah Bangsri Jepara.

Berbeda dengan lembaga lain, sekolah ini memiliki mata pelajaran khusus, yaitu mata pelajaran takhassus yang mempelajari tentang ilmu keagamaan seperti Fiqih, Tauhid, Bahasa Arab, Tajwid Al Qur'an, Nahwu, Hadist, Akhlak dan Ke-NU an. Mata pelajaran tersebut biasanya diikuti sertakan di perlombaan baik tingkat Kabupaten, Nasional maupun Internasional. Namun, dalam menentukan rekomendasi siswa berprestasi di mata pelajaran tersebut bukanlah hal yang mudah. Hal tersebut dilihat dari segi penguasaan materi pelajaran baik secara teori maupun praktik. Bukan hanya itu, nilai sikap dan daftar hadir juga penting untuk dipertimbangkan. Oleh sebab itu, pihak sekolah memerlukan sebuah metode data mining *clustering*, yaitu K-Means untuk menyelesaikan permasalahan tersebut dengan cepat, tepat dan akurat.

Clustering merupakan sebuah metode yang digunakan untuk mencari dan mengelompokkan data yang memiliki kesamaan karakteristik antara data satu dengan data lainnya. Terdapat beberapa algoritma *clustering*, salah satunya K-Means. Algoritma K-Means merupakan algoritma pengklasteran yang cukup sederhana, yaitu dengan mempartisi database kedalam beberapa klasteran [2].

2. Metode Penelitian

Desain penelitian yang digunakan pada penelitian ini adalah analisis kuantitatif. Metode tersebut menekankan analisisnya terhadap data *numeric* sebagai penunjang apa yang akan dicapai pada penelitian. Peneliti menggunakan data nilai tugas, Ujian Tengah Semester (UTS), Ujian Akhir Semester (UAS) mata pelajaran takhassus, nilai sikap dan daftar hadir siswa yang diambil dari kelas VIII semester genap tahun ajaran 2015/2016 sampai dengan 2019/2020.

3. Hasil dan Pembahasan

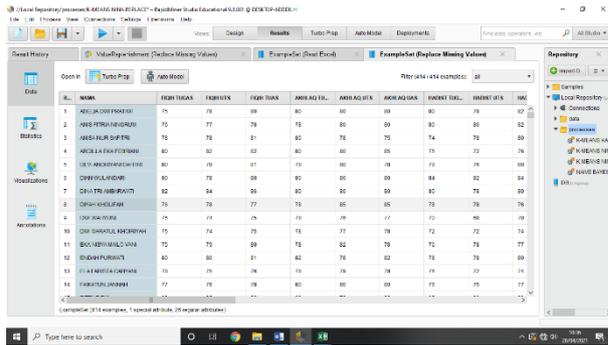
Setelah tahapan pengumpulan data dilakukan, diperoleh data sebanyak 414 yang memiliki beberapa atribut, antara lain: nama siswa, nilai tugas, Ujian Tengah Semester (UTS), Ujian Akhir Semester (UAS) dari masing-masing mata pelajaran takhassus, nilai sikap serta daftar hadir siswa yang terdiri dari sakit, ijin, dan absen. Kemudian dilakukan pengolahan data awal sebelum proses *clustering* dilakukan.

Langkah pertama, nilai di masing-masing atribut daftar hadir dijumlahkan. Perhitungan dilakukan pada *Microsoft Excel*, sehingga diperoleh hasil sebagai berikut:

NO	NAMA	FRESH TUGAS	FRESH UTS	FRESH TUGAS	AKHIR TUGAS	AKHIR UTS	AKHIR TUGAS	HABSK TUGAS	HABSK UTS	HABSK TUGAS	DAFTAR HADIR	NILAI SIKAP
1	ANJELIA DWI PRATIWI	75	78	80	80	80	80	80	78	82	2	80
2	ANGUS FERDA NINGSIDI	78	77	78	78	80	80	80	80	80	0	80
3	ANGUS NUR SAFITRI	78	78	82	80	78	75	74	78	80	0	80
4	ANJELIA EKA FEBRIANI	80	82	80	80	80	85	75	72	78	0	80
5	DELI ANDRIYANI SAFITRI	80	78	82	78	80	78	78	80	80	1	80
6	DANI WILANDARI	80	78	80	80	80	80	84	82	84	0	80
7	DONIA TER SANDIRAKATI	82	84	80	80	80	80	80	78	80	2	80
8	DIYAH KHOLIFAH	78	78	77	78	85	85	78	78	78	1	80
9	DINA HARITHA	78	78	78	78	78	77	76	86	78	1	80
10	DINA MAKATUL KHOIRINIAH	75	78	75	78	77	78	72	72	74	0	80
11	EKA NISYA MALID YANI	75	75	80	78	82	78	76	78	77	0	80
12	ENDAH PURNAMA	80	80	82	82	78	82	78	78	80	1	80
13	EKA FARISTA CAHYANI	78	75	78	78	79	79	76	72	78	1	80
14	FANITA ULIA HANIM	77	75	78	80	80	80	76	75	77	0	80
15	FETA RIDKI	80	82	82	80	78	82	82	80	80	0	85
16	HESMILA VA CAHYANI	77	78	78	78	79	80	72	78	72	0	80
17	HESMILA CAHYANI	75	78	75	78	77	77	76	73	72	0	80
18	INTAN SITI RICHANA	78	77	78	80	79	82	77	78	78	0	80
19	INDONESIA NIS SAROH	78	78	77	80	80	80	78	75	77	2	80

Gambar 1. Penjumlahan Daftar Hadir

Selanjutnya adalah pembersihan data yang dilakukan pada *software* RapidMiner Versi 9.6 untuk mengidentifikasi dan menghilangkan data yang kosong (*missing value*) di masing-masing atribut selain nama siswa. Sehingga diperoleh hasil sebagai berikut:



Gambar 2. Hasil Pembersihan Data

Pada gambar diatas terdapat keterangan filter sebanyak data yang digunakan, yaitu 414 data. Artinya, dari proses pembersihan data tidak terdeteksi adanya data yang kosong (*missing values*).

Setelah dilakukan pengolahan data awal, langkah selanjutnya adalah perhitungan K-Means menggunakan langkah-langkah sebagai berikut:

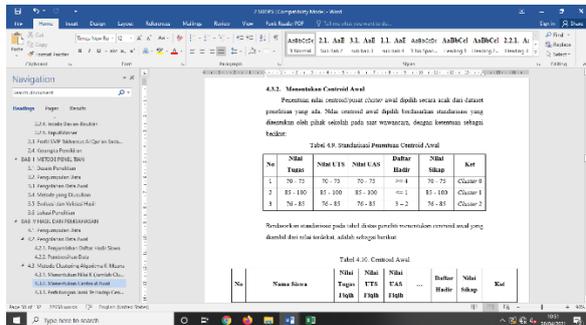
1. Menentukan nilai K (Jumlah Cluster)

Ditentukan jumlah *cluster* yang digunakan sebanyak $k=3$, terbagi kedalam *cluster 0* adalah siswa kurang berprestasi, *cluster 1* adalah siswa berprestasi, dan *cluster 2* adalah siswa berpotensi berprestasi.

2. Menentukan Centroid Awal

Nilai centroid/pusat *cluster* awal dipilih secara acak dari dataset penelitian yang ada. Nilai centroid awal dipilih berdasarkan standarisasi yang ditentukan oleh pihak sekolah pada saat wawancara, dengan ketentuan sebagai berikut:

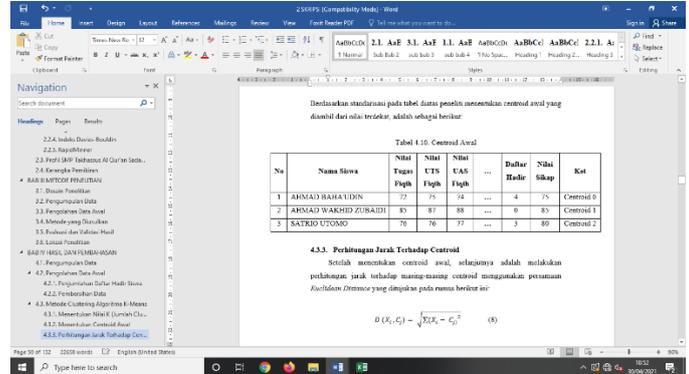
Tabel 2. Standarisasi Sekolah



Berdasarkan standarisasi pada tabel diatas peneliti menentukan centroid awal yang

diambil dari nilai terdekat, adalah sebagai berikut:

Tabel 3. Pusat Cluster Awal



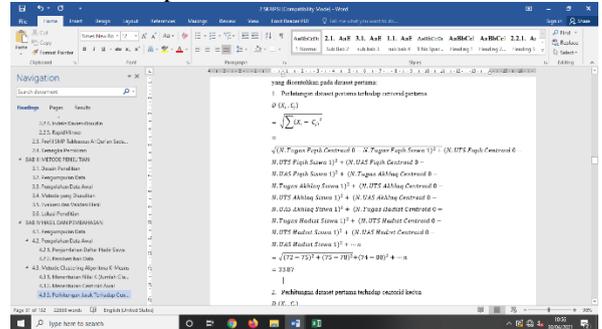
3. Perhitungan Jarak Terhadap Centroid

Menggunakan persamaan *Euclidean Distance* yang ditujukan pada rumus berikut ini:

$$D(X_i, C_j) = \sqrt{(X_i - C_j)^2} \quad (7)$$

Sebagai contoh pada dataset pertama:

1. Perhitungan dataset pertama terhadap centroid pertama



4. Pengelompokan Data

Data dikelompokkan berdasarkan kedekatan terhadap centroid dengan cara mengambil nilai terkecil dari hasil perhitungan jarak terhadap masing-masing centroid. Maka, diperoleh hasil sebagai berikut:

Tabel 4. Hasil Iterasi Pertama

No	Nama Siswa	Jarak Terlewat	Jarak Tepat	Jarak Terlambat	Jumlah Anggota	Jarak Cluster
1	ANANDA PRATIWI	20,24	14,75	9,54	9,24	0,2
2	ANNA FITRIYAH SYAHID	30,92	19,90	12,41	12,17	0,2
3	ANNA NUR KHATIRI	14,48	11,90	9,21	9,11	0,2
4	ARDELA RICA FERIANI	31,16	22,52	22,98	22,58	0,2
5	AN ANANDIYAN KHATIRI	20,70	20,94	18,87	18,62	0,2
6	DIANA SYLWANDI	20,10	22,81	18,11	18,11	0,2
7	DIANA TRI ANGGAWATI	40,24	21,76	21,09	21,09	0,2
8	DYAH KIRI RIFAN	11,74	18,90	14,48	14,48	0,2
9	DWI WANGI	1,40	21,60	17,76	7,40	0,0
10	DWI SAKAPRI KHORRIYAH	4,16	20,81	10,21	6,16	0,0
11	EKA NISYA KHALID VANI	18,20	14,24	9,71	9,71	0,2
12	INDAH FERIANI	24,91	20,91	9,21	9,21	0,2
13	ILVA FARIHA CAHYANI	13,24	22,21	6,92	6,92	0,2
14	FAKATUNNANNA	11,95	13,11	7,93	7,93	0,2
15	FERYA WATI	24,74	22,74	18,60	18,60	0,2
16	IRISYALIL CAHYANI	6,20	14,10	10,91	6,20	0,0
17	HILLMALLA CAHYANI	6,20	12,91	11,72	6,20	0,0
18	INTAN SETIYANING	30,04	17,92	7,92	7,92	0,2
19	KAROLINA SYLWANDI	14,31	20,71	12,20	12,20	0,2
20	NURIE HENDAYATI K	21,18	13,10	14,88	14,88	0,2

Berikut adalah jumlah masing-masing cluster berdasarkan hasil pengelompokan pada iterasi pertama diatas:

Tabel 5. Cluster Model Iterasi Pertama

Cluster	Jumlah
C0	56
C1	50
C2	308

5. Pengulangan Iterasi

Setelah melakukan pengolahan data pada iterasi pertama, langkah selanjutnya adalah mengulangi tahapan pengolahan data tersebut hingga memperoleh hasil pengelompokan cluster yang sama dengan iterasi yang dilakukan sebelumnya dan nilai tidak berubah lagi. Pada penelitian ini dilakukan pengulangan iterasi hingga ke delapan.

Nilai centroid baru didapat dari nilai rata-rata hasil penjumlahan seluruh anggota masing-masing cluster kemudian dibagi jumlah anggota masing-masing cluster hasil iterasi ketujuh.

Tabel 6. Centroid Baru

Nilai Tengah High	Nilai LRA High	Datar Hadir	Nilai Silap	Jumlah Anggota Cluster Iterasi 7
C0	72,194	72,071	72,899	3,090
C1	81,892	84,038	85,712	0,135
C2	77,320	77,148	78,398	0,420

Tabel 7. Hasil Iterasi Ketujuh

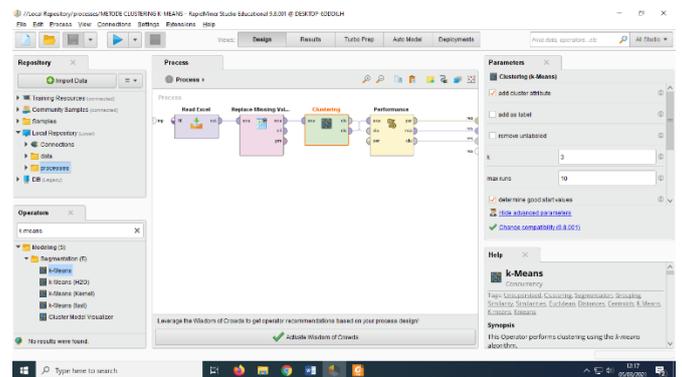
No	Nama Siswa	Jarak Terlewat	Jarak Tepat	Jarak Terlambat	Jumlah Anggota	Jarak Cluster
1	ANANDA PRATIWI	20,24	14,75	9,54	9,24	0,2
2	ANNA FITRIYAH SYAHID	30,92	19,90	12,41	12,17	0,2
3	ANNA NUR KHATIRI	14,48	11,90	9,21	9,11	0,2
4	ARDELA RICA FERIANI	31,16	22,52	22,98	22,58	0,2
5	AN ANANDIYAN KHATIRI	20,70	20,94	18,87	18,62	0,2
6	DIANA SYLWANDI	20,10	22,81	18,11	18,11	0,2
7	DIANA TRI ANGGAWATI	40,24	21,76	21,09	21,09	0,2
8	DYAH KIRI RIFAN	11,74	18,90	14,48	14,48	0,2
9	DWI WANGI	1,40	21,60	17,76	7,40	0,0
10	DWI SAKAPRI KHORRIYAH	4,16	20,81	10,21	6,16	0,0
11	EKA NISYA KHALID VANI	18,20	14,24	9,71	9,71	0,2
12	INDAH FERIANI	24,91	20,91	9,21	9,21	0,2
13	ILVA FARIHA CAHYANI	13,24	22,21	6,92	6,92	0,2
14	FAKATUNNANNA	11,95	13,11	7,93	7,93	0,2
15	FERYA WATI	24,74	22,74	18,60	18,60	0,2
16	IRISYALIL CAHYANI	6,20	14,10	10,91	6,20	0,0
17	HILLMALLA CAHYANI	6,20	12,91	11,72	6,20	0,0
18	INTAN SETIYANING	30,04	17,92	7,92	7,92	0,2
19	KAROLINA SYLWANDI	14,31	20,71	12,20	12,20	0,2
20	NURIE HENDAYATI K	21,18	13,10	14,88	14,88	0,2

Berikut adalah jumlah masing-masing cluster berdasarkan hasil pengelompokan pada iterasi kedelapan diatas:

Tabel 8. Cluster Model Iterasi Ketujuh

Cluster	Jumlah
C0	107
C1	51
C2	256

Berikut adalah permodelan pada RapidMinner:

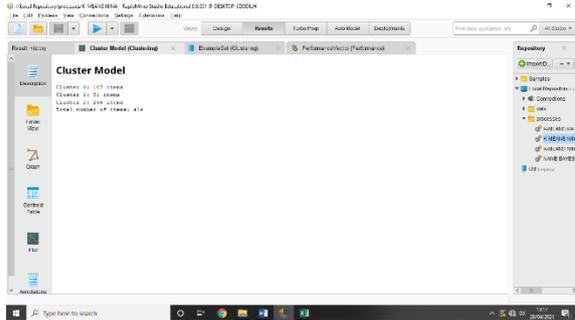


Gambar 3. Permodelan K-Means

Sehingga pada tools RapidMinner diperoleh hasil cluster sebagai berikut:

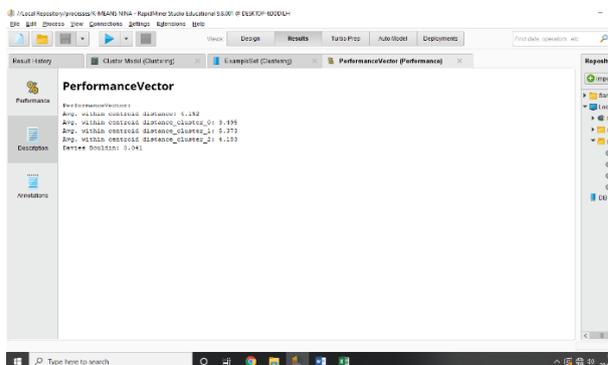
Row No.	NAMA	cluster	FIJH TUGAS	FIJH UTS	FIJH UAS	ARRAQ TIL	ARRAQ ULS	ARRAQ IAS
39	ARHO HOING WHIFLO	cluster_2	82	80	82	79	78	80
41	ARHO BAGO SYAHID	cluster_2	70	77	78	80	79	80
42	ARHO DEO RINI	cluster_3	77	78	81	76	77	76
42	ARHO FALDI	cluster_2	75	76	77	78	79	80
43	AL BANSYRI KHATIRI LAILI	cluster_3	80	78	80	80	78	79
44	ARNO ABRIYAN	cluster_0	75	76	75	76	75	76
45	ANILKA RIZQIYAH ULLE	cluster_1	79	79	80	80	81	82
46	DHAR AL ABANIBARI	cluster_2	78	78	78	80	82	84
47	DHAR SYAHID	cluster_2	70	75	70	78	79	81
48	DHAR CHORUL UMBA	cluster_1	75	75	74	78	79	78
48	DHAR RAGAS ARDANA	cluster_0	79	74	75	78	77	77
50	DHAR RINI HENDAT	cluster_2	77	77	78	78	79	78
51	DHAR RIZQIYAH DWI	cluster_2	78	78	78	77	78	78
52	DHAR PRATIWI	cluster_0	75	77	76	76	77	76

Gambar 4. Hasil Cluster RapidMinner



Gambar 5. Cluster Model RapidMiner

Setelah melakukan *clustering* pada data, selanjutnya dilakukan Pengujian validitas menggunakan Indeks *Davies-Bouldin* dan pada RapidMiner diperoleh hasil sebagai berikut:



Gambar 6. Hasil Indeks *Davies-Bouldin*

4. Kesimpulan

Berdasarkan evaluasi dan validasi hasil Indeks *Davies-Bouldin* menggunakan dataset sebanyak 414, dapat disimpulkan bahwa metode *clustering* K-Means memiliki kinerja yang cukup baik dalam mengelompokkan siswa berprestasi mata pelajaran takhassus di SMP Takhassus Al Qur'an Sadamiyyah Bangsri Jepara. Peneliti menggunakan nilai k (*cluster*) = 3 yang terbagi kedalam siswa kurang berprestasi, siswa berprestasi, dan siswa berpotensi berprestasi, sehingga memperoleh hasil dari perhitungan IDB sebesar 0,673 pada *Microsoft Excel* dan 0,041 pada RapidMiner. Sedangkan hasil pengelompokkan pada *Microsoft Excel* dan RapidMiner memperoleh hasil yang sama, yakni sebanyak 107 siswa termasuk kedalam siswa kurang berprestasi, 51 siswa termasuk

kedalam siswa berprestasi, dan sebanyak 256 siswa termasuk kedalam siswa berpotensi berprestasi.

Referensi

- [1] Poerwanto, "Evaluasi Pengajaran Siswa," Rosda Edition, 2012.
- [2] M. Z. B. B. N. Fauziah Nur, "Penerapan Algoritma K-Means Pada Siswa Baru Sekolah Menengah Kejuruan Untuk Clustering Jurusan," *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 1, no. 2, pp. 100-105, 2017.
- [3] M. T. H. S. A. Hari Witono, "Identifikasi Siswa Berprestasi di bawah Kemampuan (Underachievement) Sekolah Dasar di Kota Mataram," *Jurnal Ilmiah Pendidikan Gusu Sekolah Dasar*, vol. V, no. I, pp. 01-14, 2021.
- [4] L. T. P. d. T. E. Aronson J.E, "Decision Support System and Intellegent System - 7 th ed. Pearson Education, Inc. Dwi Prabantini (Penerjemah). Sistem Pendukung Keputusan dan Sistem Cerdas," Yogyakarta, ANDI, 2005.
- [5] A. P. d. A. M. Siregar, DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner, Surakarta: CV Kekata Group, 2017.
- [6] C. W. A. P. W. A. & I. E. Astria, "Metode K-Means Pada Pengelompokan Wilayah Pendistribusian Listrik," *Seminar Nasional Sains Dan Teknologi (SENSASI)*, vol. 2, p. 1, 2019.
- [7] D. L. D. d. D. W. Bouldin, "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2), 1979, p. 224.
- [8] A. A. G. N. F. U. R. F. R. U. Y. P. P. N. R. Brilian Rahmat, "Implemetasi K-Means Clustering pada Rapidminer untuk Analisis Daerah Rawan Kecelakaan," *Seminar Nasional Riset Kuantitatif Terapan*, pp. 58-62, 2017.