

Implementasi Metode Clustering Algoritma K-Means untuk Menentukan Kelompok Tahfidz dan Tahsin di Pesantren Siswa Al-Ma'soem

Sofia Dewi¹, Siti Fatimah²

^{1,2}Sistem Informasi, Universitas Ma'soem, Indonesia
sofiadewi.job@gmail.com

Info Artikel

Sejarah artikel :

Diterima Februari 2023

Direvisi Maret 2023

Disetujui Maret 2023

Diterbitkan Maret 2023

ABSTRACT

Determining the tahfidz and tahsin groups with computer programs is an effort that needs to be done to find out the ability of students during this tahfidz program. At this time, in the process of grouping tahfidz and tahsin at pesantren Siswa Al Ma'soem, it is necessary to Cluster with data so that the ability of students in answering the test can be known the student's self-limit and enter the tahfidz or tahsin group. In this study, 349 data were used as the main data for clustering using the K-Means algorithm with tests carried out on Rapidminer and Microsoft Excel. Based on the results of the analysis in the research conducted, it can be concluded that from 4 attributes. In Cluster 1 with a total of 177 students, it was declared into the tahfidz group. In addition, other cluster with a total of 172 students who entered the tahsin group with validation of 0.312 and 0.238 in Microsoft Excel and Rapidminer which stated that the smaller the validation number, the better the cluster used.

Keywords : Algorithm; Clustering; K-Means; Rapidminer.

ABSTRAK

Menentukan kelompok tahfidz dan tahsin dengan program komputer merupakan suatu usaha yang perlu dilakukan untuk mengetahui kemampuan santri selama melakukan program tahfidz ini. Pada saat ini dalam proses pengelompokan tahfidz dan tahsin di Pesantren Siswa Al Ma'soem perlu dilakukan pengklasterisasian dengan data agar kemampuan santri dalam menjawab tes dapat diketahui batasan diri santri tersebut dan masuk kedalam kelompok tahfidz atau tahsin. Pada penelitian ini menggunakan 349 data sebagai data utama untuk melakukan klasterisasi menggunakan algoritma *K-Means* dengan pengujian yang dilakukan pada *Rapidminer* dan Microsoft Excel. Berdasarkan hasil analisa pada penelitian ini dapat disimpulkan bahwa dari 4 atribut. Pada *cluster* 1 dengan jumlah santri 177 dinyatakan ke dalam kelompok tahfidz. Sedangkan *cluster* lain yang dengan jumlah 172 santri yang masuk kedalam kelompok tahsin dengan validasi 0,312 dan 0,238 pada Microsoft Excel dan *Rapidminer* yang menyatakan semakin kecil angka validasi semakin baik *cluster* yang digunakan.

Kata Kunci : Algoritma; Klasterisasi; K-Means; *Rapidminer*.

PENDAHULUAN

Teknologi informasi dapat mendukung dalam hal membuat, mengubah, menyimpan, mengkomunikasikan suatu informasi dan menyebarkan informasi. Dengan adanya teknologi informasi manusia dapat dipermudah dalam mendapatkan informasi secara cepat dan tepat yang salah satunya dalam bidang

pendidikan. Melalui pendidikan seseorang diharapkan mendapatkan ilmu pengetahuan yang lebih untuk meningkatkan kehidupan yang lebih sempurna. Salah satu kualitas pendidikan yang baik dilihat dari prestasi belajar seseorang yang didasarkan salah satunya dengan nilai pembelajaran pada menempuh pendidikan [1].

Adapun pendidikan berupa pesantren di Jawa Barat menurut Pangkalan Data Pondok Pesantren (2019) untuk Jawa Barat menyumbang jumlah pondok pesantren terbanyak yakni 8.410 pesantren dari total pesantren nasional, untuk Bandung sendiri berjumlah 127 pesantren. Pesantren Siswa Al Ma'soem (PSAM) salah satu pondok Pesantren yang semula bernama Pesantren Pendidikan Formal, Mulai dibangun pada tahun 2000. Program Tahfidz Al-Quran di Pesantren Siswa Al Ma'soem adalah program menghafal Al-Quran, santri dapat menghafal minimal 5 juz selama mengenyam pendidikan di sekolah dan pesantren. Untuk mengetahui kemampuan santri sebagai tumpuan dalam membuat kelompok. Hal ini menjadi penting karena sesuai dengan fakta yang ada, santri dengan kemampuan tahfidznya lancar dengan mengacu dari beberapa indikator yang menjadi penilaian hafalan, sedangkan santri yang memiliki kemampuan yang masih dasar atau kurang, akan dikelompokkan dengan kelompok yang sesuai dengan kemampuan dasar juga. Maka dari itu pengelompokan ini adalah yang paling tepat berdasarkan penilaian yang telah diambil. Dan kemampuan didalamnya terdapat kompetensi yang seimbang, sehingga dapat memotivasi santri agar lebih giat lagi dalam belajar menghafal tahfidznya.

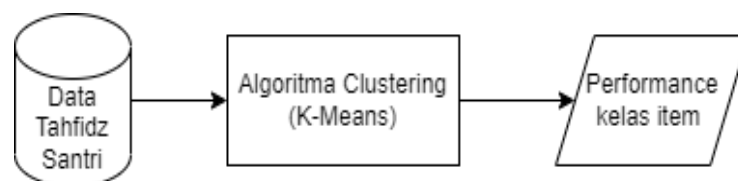
Tahfidz Al-qur'an dirasakan sangat perlu dikembangkan. Banyak lembaga pendidikan islam di Indonesia saat ini yang mengembangkan program tahfidz al-qur'an. Asal kata Tahfidz berasal dari kata *hifz* atau *hafiza*. Berdasarkan kamus Al-Munawwir kata tahfidz merupakan bentuk kata benda (masdar) dari kata haffadza yang artinya mendorong diri agar menghafalkan Al-Quran [2]. Adapun pengertian dari kata "Tahsin" secara bahasa diambil dari kata kerja *حَسَّنَ-يُحَسِّنُ-تَحْسِينًا* yang memiliki arti memperbaiki, memperbaiki atau membuat lebih baik dari semula. Tahsin dalam islam mengandung makna bahwa tuntutan agar dalam membaca Al-Quran harus benar dan tepat sesuai dengan contoh untuk menjaga keaslian tilawah sesuai dengan sunnah Rasulullah SAW [3]. Dalam proses pengelompokan tahfidz dan tahsin ini seorang Musyrif/penguji tahfidz masih menggunakan metode manual, yaitu pada penulisan tanda benar dan salah pada hasil kemampuan tes tahfidz santri.

Dalam mengatasi masalah tersebut, peneliti akan menggunakan Data Mining metode *Clustering* algoritma *K-Means* dalam rangka untuk melakukan penentuan kelompok tahfidz dan tahsin dilihat dari kemampuan yang dimiliki oleh santri. *Data mining* adalah suatu proses menganalisis pola data yang tersembunyi menurut berbagai perspektif untuk kategorisasi menjadi informasi yang berguna, yang dikumpulkan di area utama, data warehouse untuk menganalisis dengan mudah, algoritma data mining, memfasilitasi pengambilan keputusan bisnis, dan informasi lainnya. *Data mining* juga adalah bagian dari *knowledge discovery in database* (KDD) [4]-[5]. Metode yang digunakan adalah melakukan Pengelompokan data-data ke dalam sejumlah kelompok (*Cluster*) berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada

[6]. Penerapan Algoritma *K-Means* merupakan salah satu algoritma dari *clustering* yang masuk dalam kelompok *unsupervised learning* yang digunakan untuk membagi data menjadi beberapa kelompok dengan sistem partisi [7]. Algoritma *K-Means* dapat diartikan sebagai algoritma pembelajaran yang sederhana untuk menentukan suatu permasalahan pengelompokan yang bertujuan untuk meminimalkan kesalahan ganda [8]. Aplikasi yang digunakan untuk pengujian proses klasterisasi adalah *RapidMiner*. *RapidMiner* merupakan perangkat lunak yang bersifat terbuka (*opensource*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap data *mining*, *text mining* dan analisis prediksi [9]-[10].

METODE

Pada tahap penelitian ini menguraikan salah satu metode dari beberapa yang ada pada data mining yaitu metode *Clustering* dengan algoritma *K-Means*. Di dalam metode ini, ada beberapa langkah yang harus dilakukan diantaranya pengumpulan data dengan cara mengumpulkan data dan mempelajari data yang didapat dari pihak penguji Tahfidz. Dari pengumpulan data setiap bulan lebih dari 80 santri yang melakukan tes tahfidz yang mana berhasil memperoleh selama 2 semester. Data yang terkumpul sekitar 634 data, kemudian dilakukan pembersihan data dari data duplikasi yang akhirnya terdapat 349 data. Data santri yang sudah lulus jika terdapat 2 data atau lebih maka diambil salah satu pada data yang memenuhi range nilai lulus. Jika data santri yang berada dibawah *range* yang telah ditentukan terdapat 2 data atau lebih maka diambil salah satu. Gambaran proses klasterisasi pada penelitian ini seperti gambar 1 dibawah ini.



Gambar 1. Tahapan Clustering

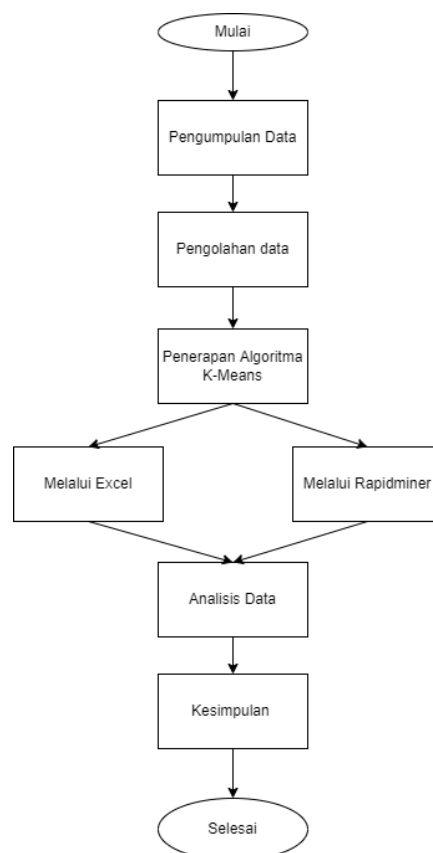
Tahapan ini adalah awal mula dari data yang sudah diperoleh, kemudian melalui perhitungan algoritma *clustering* yaitu *k-means* yang termasuk kedalam *partitional clustering*.

1. Data Tahfidz santri; didapatkan berupa lembaran kertas penilaian tahfidz santri melalui pengujian tahfidz. kemudian data ini diinputkan ke dalam excel.
2. Algoritma *Clustering*; digunakan untuk mengolah data untuk pengelompokan data dengan *clustering k-means* yang nantinya akan terbentuk *cluster-cluster* dan dilakukan analisis.
3. *Performance*; diperlukan untuk melakukan evaluasi pada tahapan algoritma *clustering* sebelumnya, *performance* ini mengirimkan sebuah list nilai kriteria pada *cluster*.

Berikut ini adalah penjelasan tahapan prosedur kerja penelitian yang dilakukan diantaranya:

1. Pengumpulan data; data yang dikumpulkan berupa data tahfidz santri Pesantren Siswa Al Ma'soem Jatinangor untuk dilakukan pengolahan data. data yang digunakan adalah data berupa nama santri, kelas, santri, wali santri, nilai tes tahfidz.
2. Pengolahan Data; semua data santri sudah ditransformasikan ke dalam bentuk angka, maka data-data ini dapat diproses ke langkah berikutnya, yaitu *data mining* menggunakan metode clustering dengan algoritma *K-Means*.
3. Penerapan Algoritma *K-Means*
 - a. Melalui Excel; penerapan pada excel dilakukan dengan cara data yang dibutuhkan akan dilakukan transformasi. karena pada penelitian ini data soal bukan berupa angka. Setelah ditransformasikan kemudian digunakan rumus perhitungan *Euclidean Distance* dilakukan beberapa kali hingga hasil klaster tidak berubah.
 - b. Melalui *Rapidminer*; penggunaan *Rapidminer* sangat sederhana, hanya diperlukan data yang akan diproses, operator yang sudah tersedia dan hasil dari klaster akan diproses dengan cepat.
4. Analisis data; diperlukan untuk menjabarkan hasil data yang telah diolah baik melalui excel ataupun melalui *Rapidminer*.
5. Kesimpulan; hasil akhir dapat disimpulkan dan diberikan kepada instansi untuk dapat menjadi masukan bagi instansi pada program tahfidz.

Gambaran dari penjelasan diatas sebagai berikut :



Gambar 2. Prosedur Kerja Penelitian

Variabel penilaian yang dilakukan dalam penelitian adalah variabel yang ditetapkan untuk diperoleh sebuah informasi mengenai hal tersebut yang kemudian dibuat kesimpulan. Adapun variabel yang digunakan dalam penelitian sebagai berikut :

Tabel 1. Tabel Indikator Penilaian

No	Indikator	Bobot Penilaian
1	Melanjutkan Ayat Berikutnya	2
2	Tebak Surat	2
3	Pertengahan Ayat	3
4	Satu Surat Penuh Pilihan	2

Berdasarkan table diatas, empat variabel tersebut terdiri dari soal yang ditanyakan kepada santri pada tes tahfidz sebagaimana akan dijelaskan sebagai berikut :

1. Melanjutkan ayat berikutnya, yaitu 2 soal melanjutkan ke ayat berikutnya baik pada surat pendek dan surat panjang.
2. Tebak surat, indikator ini meliputi 2 soal yaitu menebak surat pendek dan surat panjang
3. Pertengahan ayat, yaitu variabel yang meliputi 3 soal yaitu ayat pendek, ayat panjang, dan ayat pilihan acak sesuai dengan keterangan juz hafalan santri.
4. Satu surat penuh pilihan, yaitu variabel yang mencakup dalam melafalkan satu surat dari awal sampai akhir.

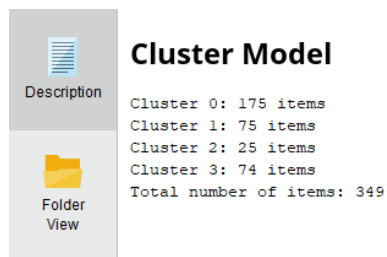
Adapun kategori pada penilaian lulus dan tidaknya santri pada tes tahfidz mengacu pada tabel 2.

Tabel 2. Kategori Penilaian

Range
0 => 8 Lulus
0 <= 8 Tidak Lulus

HASIL DAN PEMBAHASAN

Penerapan algoritma *K-Means* dengan menggunakan *Rapidminer Studio*. Aplikasi ini adalah salah satu *software* atau program yang digunakan untuk melakukan *data mining*, teks mining dan prediksi. Dengan perangkat lunak ini berbagai macam metode dan algoritma biasa digunakan sesuai dengan kebutuhan analisa. Dengan menggunakan aplikasi *Rapidminer* dalam penelitian, diperlukan persiapan data yang akan masukkan kedalam aplikasi *Rapidminer*. Penelitian ini menyiapkan data dengan menggunakan aplikasi Microsoft Excel. Penerapan/pengujian menggunakan *Rapidminer* yakni memasukkan data bisa melalui import data seperti yang tertera pada *repository* atau dengan menggunakan *Read Excel*. Dalam penelitian ini *Cluster* yang digunakan yaitu 4 *Cluster* dimulai dari *Cluster 0*, *Cluster 1*, *Cluster 2*, *Cluster 3*. Serupa dengan pendefinisian nilai pada *Cluster 0* berjumlah 175 data, *Cluster 1* berjumlah 75 data, *Cluster 2* berjumlah 25 data dan *Cluster 3* berjumlah 74 data, seperti pada gambar 3 dibawah ini.



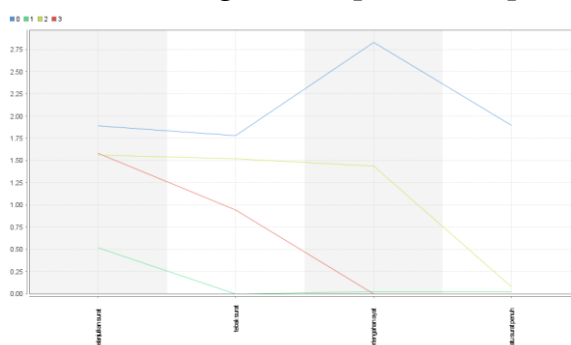
Gambar 3. Hasil Cluster Model

Dengan menggunakan *Rapidminer* hasil perhitungan jarak *Cluster* dengan *Centroid* dapat dilihat seperti pada gambar 4 dibawah ini.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
melanjutkan surat	1.891	0.520	1.560	1.581
tebak surat	1.783	0	1.520	0.946
perfengahan ayat	2.829	0.027	1.440	0
satu surat penuh	1.897	0.027	0.080	0

Gambar 4. Hasil Perhitungan Jarak *Cluster* pada *Centroid*

Hasil set *cluster* berdasarkan grafik, dapat dilihat pada gambar 5.



Gambar 5. Diagram Pencar Hasil Set Clustering

Pada penerapan algoritma *K-Means* dengan menggunakan software Microsoft Excel, penelitian ini menggunakan 4 *Cluster*. Untuk menentukan titik pusat *Cluster* yang biasa disebut dengan *Centroid* dapat dilakukan dengan menggunakan rata-rata pada setiap atribut atau dengan menggunakan metode acak. Pada penelitian ini metode yang digunakan adalah metode acak.

Tabel 3. Titik Pusat Awal *Cluster*

Zalfa Nadzifah	2	2	3	2	Cluster 1
Febrianti	2	2	1	0	Cluster 2
Fadhlan Raka	2	1	0	0	Cluster 3
Daifina Salsabila	0	0	0	0	Cluster 4

Menghitung Jarak Data ke Setiap *Cluster*

Setelah menentukan jumlah *Cluster* dan *Centroid*, selanjutnya adalah melakukan perhitungan jarak data ke setiap *Cluster*. Peneliti menggunakan

Euclidean Distance yang bisa digunakan dengan mudah, dapat dipahami dan mendukung hasil perhitungan Cluster. Rumus tersebut sebagai berikut :

=SQRT((data pertama melanjutkan surat berikutnya-Centroid melanjutkan surat berikutnya)²+(data pertama tebak surat-Centroid tebak surat)²+(data pertama pertengahan ayat - Centroid pertengahan ayat)²+(data pertama membaca surah awal akhir-Centroid membaca surah awal akhir)²).

$Cluster\ 1 = \sqrt{(2 - 2)^2 + (1 - 2)^2 + (0 - 3)^2 + (0 - 2)^2} =$
$Cluster\ 2 = \sqrt{(2 - 2)^2 + (1 - 2)^2 + (0 - 1)^2 + (0 - 0)^2} =$
$Cluster\ 3 = \sqrt{(2 - 2)^2 + (1 - 1)^2 + (0 - 0)^2 + (0 - 0)^2} =$
$Cluster\ 4 = \sqrt{(2 - 0)^2 + (1 - 0)^2 + (0 - 0)^2 + (0 - 0)^2} =$

Gambar 6. Perhitungan Jarak terhadap Centroid

Memasukkan data ke dalam Cluster. Dengan rumus excel sebagai berikut :
 =IF(data pertama Cluster1=min(data pertama Cluster1;data pertama Cluster2;data pertama Cluster3;data pertama Cluster4);1;IF(data pertama Cluster2=min(data pertama Cluster1;data pertama Cluster2;data pertama Cluster3;data pertama Cluster4);2;IF(data pertama Cluster3=min(data pertama Cluster1;data pertama Cluster2;data pertama Cluster3;data pertama Cluster4);3;4))). Langkah selanjutnya adalah menentukan titik pusat Cluster baru. Langkah ini menggunakan nilai rata-rata pada data yang telah dikelompokkan pada masing-masing Cluster.

- Cluster 1 yang memiliki jumlah 177 data

Tabel 4. Centroid Baru Cluster 1

Titik Pusat Cluster Baru	Melanjutkan surat berikutnya	Tebak surah	Pertengahan surah	Membacakan surat awal-akhir
Cluster 1	1,892655367	1,774011299	2,81920904	1,88700565

- Cluster 2 berjumlah 37 data

Tabel 5. Centroid Baru Cluster 2

Titik Pusat Cluster Baru	Melanjutkan surat berikutnya	Tebak surah	Pertengahan surah	Membacakan surat awal-akhir
Cluster 2	1,378378378	1,72972973	0,864864865	0

- Cluster 3 berjumlah 60 data

Tabel 6. Centroid Baru Cluster 3

Titik Pusat Cluster Baru	Melanjutkan surat berikutnya	Tebak surah	Pertengahan surah	Membacakan surat awal-akhir
Cluster 3	1,68333	0,7	0	0

4. Cluster 4 berjumlah 75 data

Tabel 7. Centroid Baru Cluster 4

Titik Pusat Cluster Baru	Melanjutkan surat berikutnya	Tebak surat	Pertengahan surat	Membacakan surat awal-akhir
Cluster 4	0,52	0	0,026666667	0,026666667

Menentukan SSW (Sum of Square Within Cluster)

Dari data tahfidz yang sudah teruji tidak mengalami perpindahan atau perubahan tempat maka langkah selanjutnya adalah menentukan SSW. Data yang diuji adalah semua data yaitu 349 data. Pada tahap ini *centroid* yang ada pada iterasi terakhir ini disesuaikan dengan *Cluster* data.

0,66401	1,10615	0,66399	0,53159
ssw1	ssw2	ssw3	ssw4

Gambar 7. Hasil SSW

Menentukan SSB (Sum of Square Between Cluster)

Penelitian ini menggunakan iterasi ketiga atau iterasi terakhir sebagai data uji validasi karena mempunyai kualitas *Centroid* yang stabil dan baik dari iterasi sebelumnya. Perhitungan SSB tidak memerlukan banyak data yang dipergunakan adalah data centroid saja.

Tabel 8. Nilai Centroid Iterasi Terakhir

Class	Melanjutkan Surat berikutnya	Tebak Surat	Pertengahan ayat	Bacaan satu surat awal-akhir
Cluster 1	1,8927	1,774	2,8192	1,887
Cluster 2	1,3611	1,8056	0,8333	0
Cluster 3	1,7049	0,6885	0,0328	0
Cluster 4	0,52	0	0,0267	0,0267

Perhitungan untuk SSB ini berbentuk matrik sesuai dengan dengan *centroid* yang sudah ditentukan.

Tabel 9. Hasil SSB

Centroid	1	2	3	4
1	0	2,7906	3,5427	4,0386
2	2,7906	0	1,4166	2,1492
3	3,5427	1,4166	0	1,3707
4	4,0386	2,1492	1,3707	0

Kemudian dilakukan perhitungan nilai *Davies Bouldin index* (DBI) yang didapatkan dari persamaan berikut :

$$DBI = \frac{1}{4} \times 1,249543 = 0,312386$$

Pada pengujian ini dapat diamati bahwa semakin kecil nilai SSW maka hasil *Clustering* yang didapat juga lebih baik. Secara DBI menginginkan nilai sekecil mungkin untuk menilai baiknya *Cluster* yang didapat.

Tabel 10. Perbandingan Hasil Perhitungan *Clustering* Algoritma *K-Means*

Perbandingan Perhitungan	Microsoft Excel				Rapidminer			
Banyak Iterasi	3				-			
Jumlah Cluster	C1	C2	C3	C4	C1	C2	C3	C4
	177	36	61	75	175	75	25	74
Jumlah Centroid	4				4			
Jumlah DBI	0,312				0,238			

PENUTUP

Berdasarkan hasil penelitian ini tentang menentukan kelompok tahfidz dan tahsin dengan metode *Clustering* Algoritma *K-Means* di Pesantren Al Ma'soem dengan penerapannya menggunakan *Microsoft Excel* dan *Rapidminer* pada data tahfidz santri Pesantren Al Ma'soem pada tahun 2021 didapatkan hasil bahwa dari 4 Cluster pada Cluster 1 dengan jumlah santri 177 dinyatakan ke dalam kelompok tahfidz. Selain itu dengan jumlah 172 santri yang masuk kedalam kelompok tahsin. Hasil yang didapat dari validasi yang dilakukan melalui Davies Bouldin Index (DBI) pada pengolahan data *Microsoft Excel* dengan nilai 0,312 dan 0,238 untuk *Rapidminer*. Semakin kecil nilai hasil dari klasterisasi maka semakin baik Cluster yang didapatkan. Sehingga dapat disimpulkan bahwa pengolahan data menggunakan *Rapidminer* lebih direkomendasikan karena penggunaannya sederhana, mudah diproses dengan cepat dan akurat.

DAFTAR PUSTAKA

- [1] S. A.M, *Interaksi dan Motivasi Belajar Mengajar*. Jakarta: PT.Raja Grafindo Persada, 2004.
- [2] Muhammad, *Strategi Pembelajaran Tahfidz Al-qur'an*. 2002.
- [3] Junaidi, *Tahsin Qur'an*. Bandung: Cipta pustaka Media Perintis, 2009.
- [4] M. dkk L, *DATA MINING*. Universitas Brawijaya Press, 2018.
- [5] M. Arhami and M. Nasir, *DATA MINING Algoritma dan Implementasi*. 2020.
- [6] E. Prasetyo, *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*. 2012.
- [7] Wanto, *Data Mining: Algoritma dan Implementasi*. Yayasan Kita Menulis, 2020.
- [8] & K. Singla, "Comparative Analysis & Evaluation of Euclidean Distance Function and Manhattan Distance Function Using K-Means Algorithm," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, 2012.
- [9] L. E. M.H, "SMK Analisis Kesehatan Abdurrah Menggunakan Algoritma," *J.Teknol. Dan Sist. Inf. Univrab*, vol. 2(2), pp. 220-233, 2017.
- [10] S. Dewi and O. Oktaviawati, "Penerapan Algoritma C4. 5 untuk Pemahaman Siswa SMK Pada Pelajaran Kompetensi Keahlian," *Intern. (Information Syst. Journal)*, vol. 5(2), pp. 116-125, 2022.