

Implementasi Algoritma K-Means Untuk Pengelompokan Kemampuan Siswa dalam Literasi di SMP Santi Yasa Petak

Ngakan Putu Widyasprana^{a1}, Ida Bagus Made Mahendra^{a2}

^{a1}Informatics Departement, Faculty of Math and Science, University of Udayana
South Kuta, Badung, Bali, Indonesia
¹ngakanputu39@gmail.com
²ibm.mahendra@unud.ac.id

Abstract

Literacy is the ability of Indonesian language competence on the basis of reading and writing. Apart from reading and writing other skills also include speaking, reading comprehension, and public speaking skills. In fact, literacy skills in Indonesia are quite behind that of other countries in the world. To increase literacy skills, we use the SAC (All Smart Children) approach with TaRL (Teaching at Right Level) learning. This approach can be realized by using one of the clustering methods, namely the K-Means Algorithm. The grouping will use data from the implementation of the literacy program at SMP Santi Yasa Petak within a span of 3 months and modeling using the Rapid Miner application. The results obtained divided the students into 2 clusters, namely the cluster of students who were able to carry out the literacy program as many as 8 people and the cluster of students who had not been able to carry out the literacy program as many as 5 people.

Keywords: Literacy, SAC (Semua Anak Cerdas), TaRL, K-Means, Rapid Miner.

1. Pendahuluan

Literasi adalah kemampuan dalam memperluas kompetensi berbahasa Indonesia dalam berbagai tujuan, khususnya yang berkaitan dengan membaca dan menulis [1]. Di Indonesia sendiri kemampuan literasi saat ini masih jauh tertinggal dari bangsa lain di dunia, permasalahan literasi merupakan salah satu masalah yang harus mendapatkan perhatian khusus oleh bangsa Indonesia [2]. Tuntutan pendidikan di era saat ini mengharuskan menghasilkan generasi muda yang memiliki empat kompetensi utama yakni kompetensi berpikir, kompetensi bekerja, kompetensi berkehidupan, dan kompetensi menggunakan teknologi [3].

Demi meningkatkan kemampuan literasi perlu adanya perubahan skema pembelajaran yang dikelompokkan berdasarkan jenis usia mereka, seperti anak umur 14 masuk ke kelas 7 SMP. Melalui pendekatan SAC (Semua Anak Cerdas) yang menerapkan pembelajaran TaRL (*Teaching at the Right Level*) memungkinkan siswa mendapatkan pembelajaran mereka sesuai dengan kemampuan tidak memperhatikan usia mereka [4]. Model pembelajaran seperti ini sudah terbukti efektif dalam mengurangi anak yang rendah literasi di berbagai negara [5].

K-Means merupakan algoritma clustering yang termasuk dalam *unsupervised learning group* yang digunakan untuk mengelompokkan data menjadi beberapa kelompok dengan sistem partisi. Algoritma ini dapat menerima data yang label kelasnya tidak diketahui dan kemudian mengelompokkannya. Algoritma K-Means dapat terdiri dari beberapa cluster yang memiliki titik pusat yang disebut Centroid. Pada penerapan algoritma K-Means, input yang diterima dapat berupa data dan jumlah cluster yang diinginkan [6].

Dalam penelitian sebelumnya oleh (Asroni et al, 2015) menggunakan algoritma K-means untuk mengklasifikasi data mahasiswa berdasarkan nilai akademiknya. Algoritma K-means membagi data menjadi 4 kluster berdasarkan dengan *class* Indeks Prestasi Kumulatif (IPK). Terlihat dari hasil kluster, terpilih cluster 1 yaitu mahasiswa dengan IPK = 3.4143, sebanyak 28 Mahasiswa dari 124 Mahasiswa (23%).

1.1. Klastering

Klastering adalah suatu metode pengelompokan data yang dilihat dari kemiripan atau kedekatannya. *Cluster* memiliki arti yang berbeda dengan *group*. Kelompok adalah kondisi yang serupa sedangkan klaster tidak harus serupa, tetapi pengelompokannya didasarkan pada kesamaan sampel yang ada. Misalnya, menggunakan rumus jarak Euclidean. Jarak Euclidean adalah yang terpendek dalam kumpulan data ke pusat massa. Clustering termasuk unsupervised learning, yaitu proses membagi sekelompok himpunan data menjadi klaster-klaster berdasarkan kesamaan berbagai nilai atribut dari *dataset* tersebut. Hal ini bertujuan untuk mempercepat waktu komputasi dengan memperoleh *cluster* yang berkualitas. *Cluster* juga dapat diartikan sebagai sekumpulan objek data dalam klaster yang sama yang mirip satu sama lain dan terpisah dari objek klaster lainnya [6].

1.2. Algoritma K-Means

Algoritma K-means clustering merupakan metode analisis data atau metode *data mining* yang melakukan proses pemodelan tanpa pengawasan (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Dua jenis data clustering yang sering digunakan dalam proses pengelompokan data, yaitu *hierarchical* dan *non-hierarchical*, dan k-means merupakan metode clustering data *non-hierarchical* atau *partitional clustering*. Dalam mengklasifikasikan data dibagi menjadi beberapa kelompok dengan beberapa *cluster*. Data diseleksi menjadi beberapa kelompok dengan kriteria yang telah ditentukan kemudian dikumpulkan menjadi satu *cluster*, dimana setiap *cluster* memiliki titik pusat yang disebut Centroid. Berikut langkah-langkah untuk melakukan optimasi menggunakan algoritma K-Means [7].

- 1) Tentukan berapa banyak *cluster* k dari dataset yang akan dibagi.
- 2) Tetapkan secara acak data k menjadi pusat awal lokasi klaster.
- 3) Untuk masing-masing data, temukan pusat *cluster* terdekat. Dengan demikian berarti masing-masing pusat *cluster* memiliki sebuah *subset* dari *dataset*, sehingga mewakili bagian dari himpunan data. Oleh karena itu, telah terbentuk *cluster* k: C1, C2, C3, ..., Cn.
Rumus :

$$J = \sum_{i=1}^k \sum_{j=1}^n \|x_j - c_i\|^2$$

Penjelasan :

J = Jarak antar data dan centroid

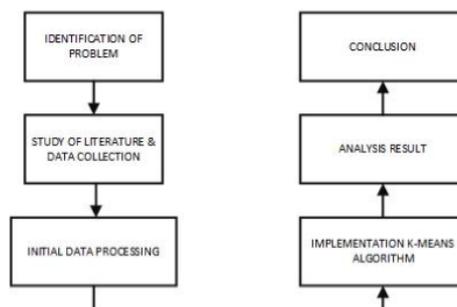
X_j^i = nilai j^{th} ke data pada cluster ke- i^{th}

C_i = nilai centroid ke- i^{th} dalam cluster ke- i^{th}

- 4) Untuk masing-masing *cluster* k, temukan pusat luasan klaster, dan perbaharui lokasi dai masing-masing pusat *cluster* ke nilai baru dari pusat latihan.
- 5) Ulangi langkah ke-3 dan ke-5 data-data pada setiap *cluster* menjadi terpusat atau selesai.

2. Metode Penelitian

Metodologi yang digunakan pada riset kali ini adalah *simulation-based*. Untuk tahapan metode penelitian dapat dilihat pada Gambar 1 .



Gambar 1. Flowchart dari Tahapan Penelitian

Dilihat dari gambar 1 diatas, proses metode penelitian yang akan terlaksana dimulai dari Identifikasi masalah (*Identification of Problem*), dilanjutkan ke proses pembelajaran literatur dan pengumpulan data (*Study of Literature and Data Collection*), tahapan pemrosesan data (*Initial Data Processing*), implementasi algoritma k-means (*Implementation K-Means Algorithm*), hasil analisis (*Analysis Result*), dan kesimpulan (*Conclusion*). Berikut penjelasan lebih lanjut tiap-tiap prosesnya :

2.1 Identifikasi Masalah

Dalam peneltian kali ini mengambil data siswa dalam kemahiran siswa membaca bahan bacaan (L1), kemampuan siswa menulis (L2), kemampuan siswa berbicara (L3), dan kemampuan menceritakan kembali (L4). Poin-poin Ini menunjukkan kemahiran siswa dalam melaksanakan literasi bahan teks bacaan.

2.2 Pembelajaran Literatur & Pengumpulan Data

Dalam tahap ini, mempelajari lebih dalam mengenai pengetahuan dalam mengumpulkan data dan cara pengambilan data sebagai bahan daftar nilai kemampuan siswa dalam literasi di SMP Santi Yasa Petak. Data siswa berasal dari hasil pelaksanaan program-program literasi di SMP Santi Yasa Petak selama program Kampus Mengajar. Pengambilan data dimulai dari tanggal 1 Agustus 2022 - 1 Oktober 2022 (3 Bulan). Dalam sebulan melaksanakan kegiatan literasi selama 8 kali, jadi total pelaksanaan sebanyak 24 kali. Untuk hasil pembelajaran literatur dan pengumpulan data dapat dilihat pada tabel 1.

Tabel 1. Total Pemetaan Kemampuan Literasi Siswa Selama 3 Bulan

Indeks	Nama Siswa	L1	L2	L3	L4
1	Anak Agung Gede Depa Sedana	1689	1660	1704	1698
2	I Gede Agus Yudi Artana	1618	1681	1651	1667
3	I Ketut Pratia Hara	1678	1704	1644	1661
4	I Putu Aditya Pratama	1647	1724	1639	1657
5	I Putu Andika	1648	1674	1710	1687
6	I Putu Juniantara	1679	1692	1689	1693
7	Ni Luh Agustina	1678	1723	1677	1687
8	Ni Luh Putu Febriani	1687	1700	1691	1666
9	Ni Made Tiarani	1662	1671	1709	1598
10	Ogi Sagita	1735	1694	1685	1696
11	Putu Aris Febriantara	1704	1642	1721	1689
12	Putu Sumerdiana	1702	1684	1690	1670
13	Rian Saputra	1654	1691	1665	1657

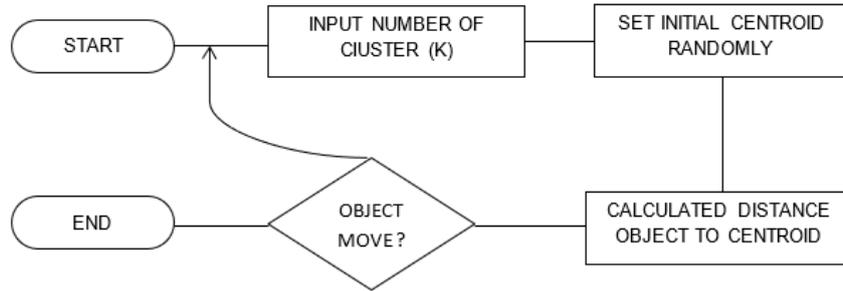
Dapat dilihat pada tabel 1, nilai siswa dari indikator pertama L1 (Kemahiran Membaca) hingga indikator keempat L4 (Kemampuan Menceritakan Kembali) sangat bervariasi dimulai dari nilai terkecil untuk indikator L1 yaitu 1618 poin hingga nilai tertinggi di 1689 poin.

2.3 Tahapan Pemrosesan Data

Dalam tahap ini, data yang didapatkan dari hasil pengamatan selama 3 bulan di sekolah SMP Santi Yasa Petak akan diproses menggunakan Algoritma K-Means. Data akan dikonversikan ke dalam bentuk integer sehingga dapat diproses oleh program. Program yang digunakan untuk melakukan klusterisasi adalah Rapid Miner untuk menunjukkan kelompok kluster dari data.

2.4 Implementasi Algoritma K-Means

Untuk proses *flowchart* implementasi algoritma k-means dapat dilihat pada gambar 2.



Gambar 2. Flowchart Algoritma K-Means

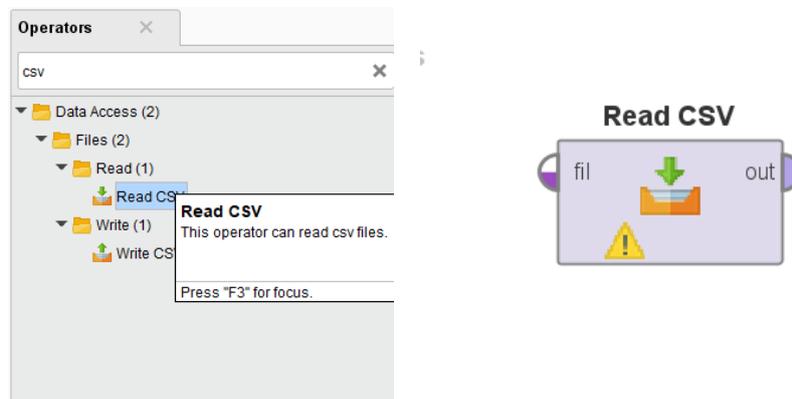
Pada gambar 2, implementasi algoritma k-means dimulai dari memasukkan jumlah kluster yang kita inginkan, dilanjutkan dengan kita menetapkan nilai centroid secara acak. Setelah itu, proses selanjutnya adalah menghitung jarak antar objek kepada centroid, jika objek masih bisa bergerak ke posisi lainnya maka akan kembali ke penentuan jumlah kluster awal. Hal ini berulang sampai ditemukan solusi yang optimal.

3. Hasil dan Pembahasan

3.1. Proses Penghitungan dengan Aplikasi Rapid Miner

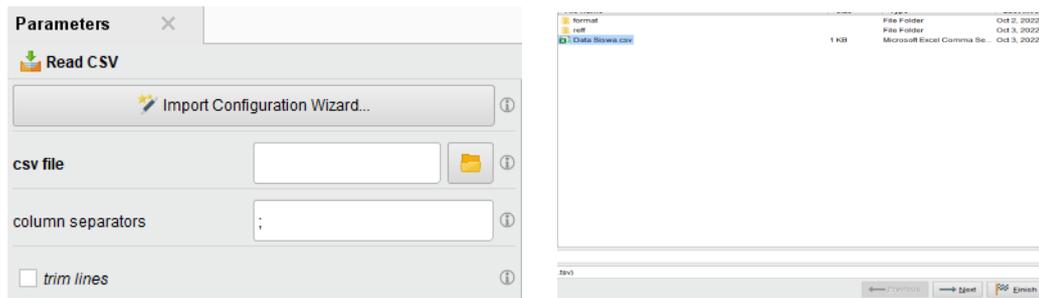
Pada tahap ini, pengimplementasian dari algoritma K-Means menggunakan aplikasi Rapid Miner, sebagai berikut :

1. Masukkan data ke dalam aplikasi Rapid Miner dalam bentuk .csv, dengan menuju menu operator dan ketikkan csv. Lalu klik 2 kali sehingga modul sudah siap untuk di masukkan data. Untuk proses lebih jelasnya dalam dilihat pada gambar 3.



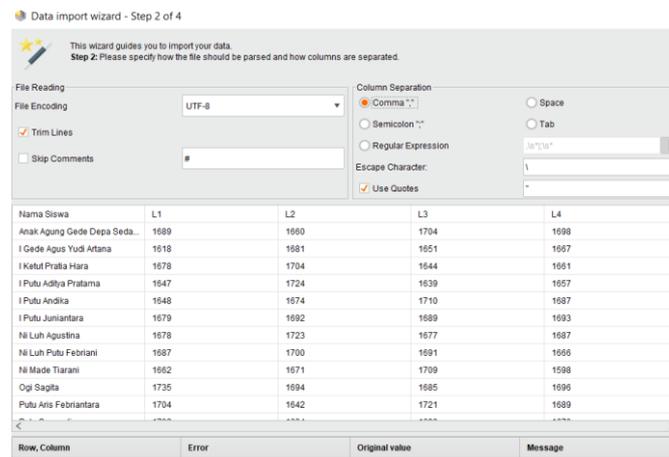
Gambar 3. Input Modul Read CSV ke Rapidminer

2. Tekan Read CSV , lalu lihat pada tab sebelah kanan akan ada opsi konfigurasi. Pilih opsi tersebut dan mulai untuk memasukkan file lalu tekan tombol selanjutnya (Next). Untuk proses lebih jelasnya dapat dilihat pada gambar 4.



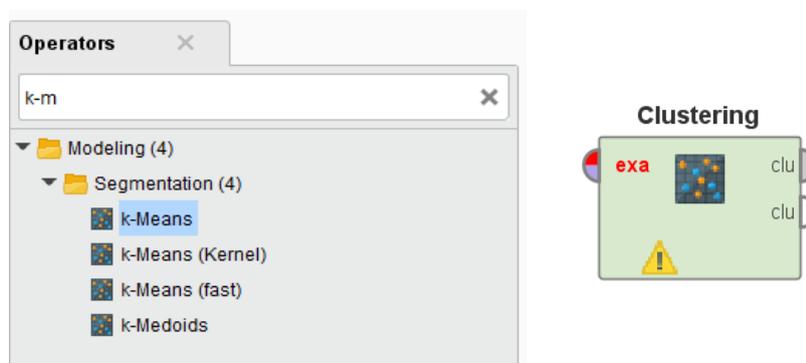
Gambar 4. Melakukan Import Data .csv

3. Setelah menekan tombol next, maka akan dibawa ke halaman untuk mengatur pemisahan data dan pengaturan lainnya. Untuk menu pilihan sesuaikan dengan format file .csv, mengenai separatornya, file encoding, dan lainnya. Lalu tekan tombol finish. Untuk lebih jelasnya dapat dilihat pada gambar 5.



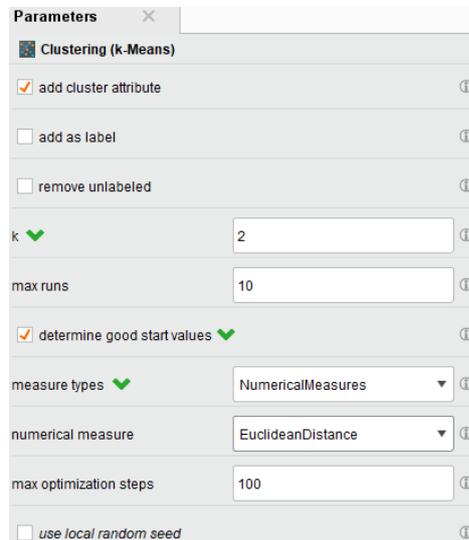
Gambar 5. Konfigurasi Data

4. Selanjutnya, memilih jenis pemrosesan data, pilih model K-Means. Dan lakukan hal yang sama seperti modul Read CSV sebelumnya (Gambar 3). Jika sudah tampil, selanjutnya tekan Clustering dan lakukan konfigurasi parameter pada tab sebelah kanan. Untuk lebih jelasnya dapat dilihat pada gambar 6.



Gambar 6. Input Modul K-Means

5. Pada menu parameter masukan 2 klaster. Pemilihan klaster ini karena *cluster_0* untuk siswa yang memiliki kemampuan literasi, dan *cluster_1* untuk siswa yang masih kurang dalam kemampuan literasi. Centang parameter "determines good start values" agar hasil klasterisasi lebih maksimal. Lalu untuk paramter lainnya adalah *default*. Untuk lebih jelasnya dapat dilihat pada gambar 7.



Gambar 7. Konfigurasi Parameter K-Means

6. Pilih Desain tab, lalu hubungkan data ke model K-Means, dan dari model K-Means hubungkan ke proses. Untuk lebih jelasnya dapat dilihat pada gambar 8.



Gambar 8. Menghubungkan Semua Modul

7. Tekan tombol run, dan hasilnya sebagai berikut :
 - A. Tampilan Tulisan
Untuk hasil klasterisasi dengan tampilan tulisan dapat dilihat pada gambar 9. Dapat dilihat hasil tampilan menghasilkan 2 klaster, dengan klaster 0 berjumlah 8 orang dan klaster 1 berjumlah 5 orang. Total data adalah 13 orang.

Cluster Model

```
Cluster 0: 8 items  
Cluster 1: 5 items  
Total number of items: 13
```

Gambar 9. Tampilan Tulisan

B. Tampilan Tabel

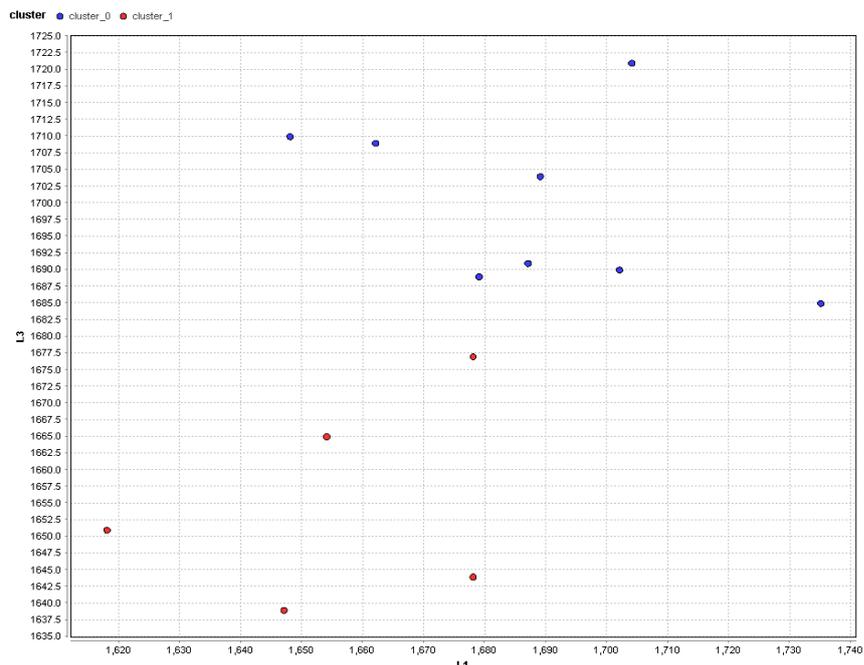
Untuk hasil klasterisasi dalam bentuk tabel dapat dilihat pada gambar 10. Dari hasil dapat dilihat nama siswa dengan jenis klaster yang mereka dapatkan. Seperti Ogi Sagita dengan klaster 0, I Putu Aditya dengan klaster 1.

id	cluster ↑	Nama Siswa	L1	L2	L3	L4
1	cluster_0	Anak Agung ...	1689	1660	1704	1698
5	cluster_0	I Putu Andika...	1648	1674	1710	1687
6	cluster_0	I Putu Juniant...	1679	1692	1689	1693
8	cluster_0	Ni Luh Putu F...	1687	1700	1691	1666
9	cluster_0	Ni Made Tiar...	1662	1671	1709	1598
10	cluster_0	Ogi Sagita	1735	1694	1685	1696
11	cluster_0	Putu Aris Feb...	1704	1642	1721	1689
12	cluster_0	Putu Sumerdi...	1702	1684	1690	1670
2	cluster_1	I Gede Agus ...	1618	1681	1651	1667
3	cluster_1	I Ketut Pratia ...	1678	1704	1644	1661
4	cluster_1	I Putu Aditya ...	1647	1724	1639	1657
7	cluster_1	Ni Luh Agustl...	1678	1723	1677	1687
13	cluster_1	Rian Saputra	1654	1691	1665	1657

Gambar 10. Tampilan Tabel

C. Tampilan Grafik Model Plot

Untuk tampilan dengan grafik model plot dapat dilihat pada gambar 11. Dilihat dari gambar terlihat warna merah adalah klaster 1, sedangkan warna biru adalah klaster 0. Nilai dari klaster 0 lebih tinggi dibandingkan dengan klaster 1.



Gambar 11. Tampilan Grafik (Plot)

4. Kesimpulan

Dari hasil penelitian terlihat bahwa pemetaan klasterisasi siswa terbagi menjadi 2 klaster. Pertama, klaster siswa yang mampu dalam melaksanakan program literasi sejumlah 8 orang (*cluster_0*) dan siswa yang belum mampu melaksanakan program literasi sejumlah 5 orang (*cluster_1*). Dengan ini dapat dikatakan bahwa siswa yang mampu melaksanakan program literasi yang di rancang di SMP Santi Yasa Petak lebih banyak dari yang tidak mampu melakukannya.

Referensi

- [1] Kementerian Pendidikan dan Kebudayaan. Panduan Gerakan Literasi Sekolah di Sekolah Dasar. Jakarta : Kemendikbud, 2016.
- [2] S. Nirmala, "Problematika Rendahnya Kemampuan Literasi Siswa Di Sekolah Dasar" *PRIMARY*, vol. 11, no. 2, p. 393-394, 2022.
- [3] Rahman. "Multiliterasi dan Pendidikan Karakter" in seminar 2nd Internasional Multiliteracy Conference and Workshop for Students and Teachers, Bandung, 2017, pp. 331-336.
- [4] M. Erfan, M. Maulidya, L. Affandi, A. Rosyidah, I. Oktaviyanti, I. Hamdani. "Identifikasi Wawasan Literasi Dasar Guru dalam Pembelajaran Berbasis Level Kemampuan Siswa" *DIDIKA*, vol. 7, no. 1, p. 1-18, 2021.
- [5] R. Banerji, M. Chavan. "Improving Literacy And Math Instruction At Scale In Indiaâ€™S Primary Schools: The Case Of Prathamâ€™S Read India Program" *Journal of Educational Change*, vol. 17, no. 4, 453-475, 2016.
- [6] Asroni, R. Andrian. "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang" *Jurnal Ilmiah Semesta Teknik*, vol. 18, no.1, p.76-82, 2015.
- [7] J. Han. M. Kamber and J. Pei, *Data Mining Concept and Techniques*, Third Edition., Waltham: Elsevier Inch, 2012.