

IMPLEMENTASI ALGORITMA K-MEANS UNTUK PEMETAAN PRESTASI AKADEMIK SISWA DI SEKOLAH DASAR TERANG BAGI BANGSA PATI

¹Yani Prihati, ²Suwarno, ³Alexander Dharmawan

^{1,2,3}Fakultas Teknik dan Informatika Universitas AKI Semarang

Jl. Imam Bonjol No. 15-17, Semarang 50173, Jawa Tengah

yani.prihati@unaki.ac.id, suwarno@unaki.ac.id, alexander.dharmawan@unaki.ac.id

Abstrak

Dalam sebuah institusi pendidikan kualitas akademik masing-masing sekolah diukur berdasarkan prestasi akademik masing-masing siswa baik itu nilai akademik setiap semester maupun nilai Ujian Nasional. Sekolah Dasar Terang Bagi Bangsa Pati merupakan salah satu sekolah yang ada di kota Pati. Seiring dengan terus bertambahnya jumlah data siswa setiap tahun, maka jumlah data yang siswa yang terus meningkat sehingga banyak penumpukan data yang belum diolah dengan optimal. Untuk itu diperlukanlah penelitian ini untuk menggali informasi dan pengetahuan baru melalui pola-pola yang terbentuk dari penumpukan data tersebut. Data mining dapat mengambil informasi yang penting dari database yang besar dan dibutuhkan manusia. Metode data mining yang biasa digunakan untuk mengcluster yaitu K-Means. Tujuan dari K-Means adalah mengelompokkan data dengan memaksimalkan kemiripan data dalam satu kluster dan meminimalkan kemiripan data antar kluster. Dengan menggunakan Metode Algoritma K-Means Clustering, dapat menentukan pengelompokan prestasi siswa tinggi, menengah dan rendah. Atribut yang digunakan pada penelitian ini yaitu 15 atribut dan menghasilkan 3 cluster. Cluster_0 rendah terdapat 2 items, cluster_1 menengah terdapat 6 items dan cluster_2 tinggi terdapat 13 items.

Kata kunci: Clustering, Data Mining, Pengelompokan prestasi siswa, K-Means

PENDAHULUAN

Sekolah merupakan media yang dijadikan oleh para peserta didik untuk menempuh pendidikannya. Setiap sekolah seharusnya dapat memantau para peserta didiknya agar dapat mengarahkan peserta didiknya untuk menjadi lebih baik. Tingginya tingkat keberhasilan siswa dan rendahnya tingkat kegagalan siswa merupakan cermin kualitas dunia pendidikan. Dunia pendidikan saat ini dituntut untuk memiliki kemampuan bersaing dengan memanfaatkan semua sumber daya yang dimiliki.

Tujuan pendidikan itu adalah mencapai standar pendidikan nasional yang meliputi standar isi, proses, kompetensi lulusan, tenaga kependidikan, sarana dan prasarana, pengelolaan, pembiayaan dan penilaian pendidikan.

Dalam sebuah institusi pendidikan kualitas akademik masing-masing sekolah diukur berdasarkan prestasi nilai akademik masing-masing siswa baik itu nilai akademik setiap semester maupun nilai ujian sekolah.

Sistem informasi dapat digunakan untuk menunjang kegiatan operasional sehari-hari sekaligus menunjang kegiatan pengambilan keputusan strategis di sekolah.

Data mining merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan

atau menemukan pola dari suatu data, sehingga data mining sering juga disebut *Knowledge Discovery in Database (KDD)* [1]. Istilah data mining mulai dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang

akademik, bisnis hingga medis.

Munculnya *data mining* didasarkan pada jumlah data yang tersimpan dalam basis data semakin besar. Dalam berbagai literatur, teori-teori pada data mining sudah ada sejak lama seperti antara lain *Naïve-Bayes* dan *Nearest Neighbour*, Pohon Keputusan, aturan asosiasi, *K-Means Clustering* dan *text mining*. [2].

Algoritma *K-Means* merupakan algoritma klusterisasi yang mengelompokkan data berdasarkan titik pusat kluster (*centroid*) terdekat dengan data. Tujuan dari *K-Means* adalah pengelompokkan data dengan memaksimalkan kemiripan data dalam satu kluster dan meminimalkan kemiripan data antar kluster. Ukuran kemiripan yang digunakan dalam kluster adalah fungsi jarak. Sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data terhadap titik *centroid* [3]. Dengan menggunakan Metode Algoritma *K-Means Clustering*, dapat menentukan pengelompokan prestasi siswa tinggi, menengah dan rendah.

Sekolah Dasar Terang Bagi Bangsa Pati merupakan salah satu sekolah yang ada di kota Pati. Secara umum penilaian prestasi siswa yaitu berdasarkan nilai pengetahuannya, baik secara teori maupun secara praktek. Serta penilaian terhadap kehadiran absensi siswa.

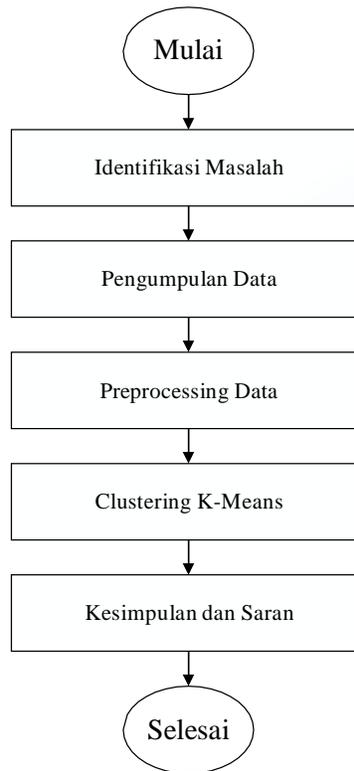
Evaluasi dan penilaian terhadap prestasi siswa dilakukan dengan memberi nilai oleh pengajar kepada semua siswa yang mengikuti pelajaran yang diajarkan dan ekstrakurikuler yang diikutinya. Seiring dengan selalu bertambahnya jumlah data siswa pada setiap tahun, maka jumlah data tersebut menyebabkan penumpukan data dan belum diolah secara optimal.

untuk itu diperlukanlah penelitian ini untuk menggali informasi dan pengetahuan baru melalui pola-pola yang terbentuk dari penumpukan data tersebut.

METODOLOGI PENELITIAN

Metodologi penelitian adalah cara yang digunakan dalam memperoleh berbagai data untuk diproses menjadi informasi yang lebih akurat sesuai permasalahan yang akan diteliti. Metodologi penelitian digunakan sebagai pedoman dalam pelaksanaan penelitian agar hasil yang dicapai tidak menyimpang dari tujuan yang telah dilakukan sebelumnya.

Berikut ini adalah metodologi yang digunakan dalam penelitian yang berjudul “Implementasi Algoritma *K-Means* untuk Pemetaan Prestasi Akademik Siswa di Sekolah Dasar Terang Bagi Bangsa Pati” dapat dilihat pada Gambar 3.1.



Gambar 1. Tahapan Metodologi Penelitian [4]

Identifikasi Masalah

Identifikasi masalah dalam penelitian ini dilakukan dengan cara wawancara. Wawancara dilakukan kepada kepala sekolah pada Sekolah Dasar Terang Bagi Bangsa Pati. Wawancara dilakukan terkait dengan masalah penumpukan data siswa yang tidak diolah dengan optimal. Untuk itu diperlukanlah penelitian ini untuk menggali informasi dan pengetahuan baru melalui pola-pola yang terbentuk dari penumpukan data tersebut. Untuk mencapai tujuan yang akan ditentukan, maka perlu dipelajari beberapa literatur yang digunakan. Kemudian literatur-literatur yang dipelajari tersebut diseleksi untuk dapat ditentukan literatur mana yang akan digunakan dalam penelitian. Melalui studi literatur, dipelajari teori-teori yang berhubungan dengan teknik *clustering* pada algoritma *K-Means* yang akan dibahas. Teori-teori tersebut diatas berdasarkan sumber berupa buku, jurnal maupun situs internet yang berhubungan dengan *clustering K-Means*.

Pengumpulan Data

Tahap pengumpulan data merupakan tahap kedua yang mana ada beberapa tahap yang akan di kerjakan mulai dari yang pertama melakukan studi pustaka dan teknik untuk memperoleh data dan adapun penjelasan sebagai berikut:

1. Studi Pustaka

Studi pustaka dilakukan dengan tujuan untuk mengetahui metode apa yang akan digunakan untuk menyelesaikan permasalahan yang akan diteliti, serta mendapatkan dasar-dasar referensi yang kuat dalam menerapkan suatu metode yang akan digunakan dalam penelitian ini, yaitu dengan mempelajari buku-buku, artiker-artikel dan jurnal-jurnal yang berhubungan dengan permasalahan yang akan dibahas.

2. Wawancara

Dalam kegiatan penelitian ini melakukan wawancara dengan kepala sekolah dan guru pada Sekolah Dasar Terang Bagi Bangsa Pati.

3. Observasi

Melakukan pengamatan secara langsung Sekolah Dasar Terang Bagi Bangsa Pati. untuk lebih mengetahui permasalahan yang diteliti dari kondisi lapangan.

Preprocessing Data

Tahapan sebelum pemrosesan data dalam *data mining* lebih dikenal dengan tahap *Preprocessing* data, adapun manfaat dari proses ini adalah untuk meningkatkan kualitas hasil pengelompokan data menggunakan algoritma *K-Means*. Penjelasan tahapan *preprocessing* data sebagai berikut:

1. Seleksi Atribut

Pada data nilai semester kelas 5 yang telah didapatkan menunjukkan jumlah atribut yang terkumpul. Selanjutnya dilakukan seleksi atribut dengan memilah atribut yang sesuai pada kebutuhan penelitian yang bertujuan untuk memperkecil ruang lingkup penelitian. Berikut merupakan 15 atribut yang digunakan dalam penelitian ini yaitu: Nama, ppkn, b.indo, ipa, ips, sbdp, mtk, b.jw, ssd, pjok, agama, b.ing, uts, uas dan absen.

2. Pembersihan Data (*Data Cleaning*)

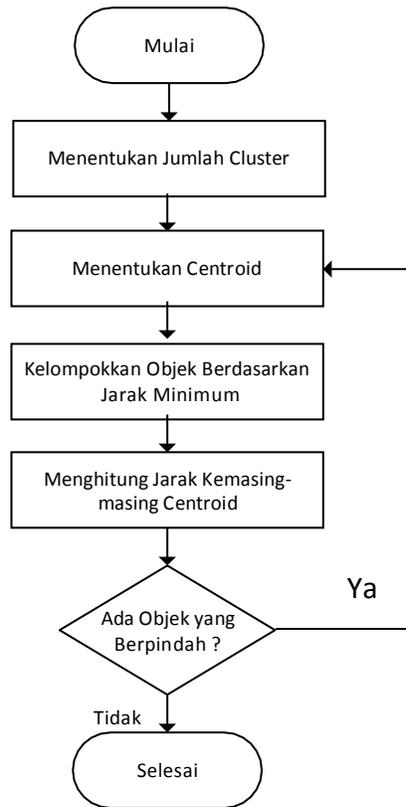
Pembersihan dilakukan untuk membuang data yang tidak layak untuk memasuki proses *data mining*, seperti data *noise* dan *missing value*. Dalam penelitian ini juga dilakukan penghapusan data yang tidak sesuai dalam penelitian seperti data yang berada diluar studi kasus penelitian.

3. Transformasi Data

Pada tahap ini dilakukan perubahan data yang memiliki tipe data yang awalnya tidak bisa diolah secara matematis menjadi data yang bisa diolah. Tujuan dari transformasi data untuk menghindari data yang rusak dan tidak valid. Pada proses *K-Means* dilakukan transformasi berupa data alpha *numerik* (teks) menjadi *numerik*.

Clustering K-Means

Pada tahap ini dilakukan penyelesaian dari algoritma *K-Means Clustering* menggunakan rumus yang telah ditetapkan. Sebagaimana yang dijelaskan pada Bab Pendahuluan bahwa pada penelitian ini bertujuan mengelompokkan menentukan prestasi siswa tinggi, menengah dan rendah, Menggunakan *K-Means Clustering*. Cara kerjanya yaitu membagi data kedalam bentuk *cluster* melalui proses tersistematis, setelah didapatkan *cluster* kemudian dilakukan analisa terhadap pola pembentukan *cluster* yang dapat menghasilkan informasi yang bermanfaat. Proses perhitungan data secara keseluruhan menggunakan *software rapidminer 7.1* untuk mengurangi dampak *human error* yang terjadi saat perhitungan manual. Adapun *flowchart* algoritma *K-Means* dapat dilihat pada Gambar 2.



Gambar 2. Proses K-Means Clustering [5]

Kesimpulan dan Saran

Tahap ini berisikan tentang kesimpulan penelitian ini dan hasil yang didapatkan. Tahap ini juga berisikan hal yang disimpulkan dan disarankan untuk melakukan pengembangan terhadap penelitian ini kedepannya.

HASIL DAN PEMBAHASAN

1. Analisa Pendahuluan

Analisa pendahuluan merupakan proses yang penting dalam sebuah penelitian. Proses analisis ini adalah proses untuk mengidentifikasi dan mengetahui hal-hal yang dilakukan sebelum akhirnya mendapat hasil dari sebuah penelitian. Analisis pendahuluan yang akan dijelaskan Analisis Studi Kasus, Analisis Metode dan Analisis Metode Terhadap Studi Kasus.

2. Pengumpulan Data

Pengumpulan data dilakukan dengan melakukan wawancara dengan kepala Sekolah Dasar Terang Bagi Bangsa Pati. Sumber data utama yaitu berasal dari data siswa kelas 5 semester 2 yang diperoleh langsung dari pihak sekolah tahun ajaran 2020/2021.

3. Preprocessing Data

Tahapan-tahapan serta proses yang dilakukan pada saat melakukan *data mining* dan proses untuk menemukan *knowledge* dapat dilihat pada sub-bab berikut ini:

a) Data Selection

Pada tahap ini dilakukan pemilihan atribut data yang akan digunakan selama penelitian. Atribut yang akan digunakan yang berhubungan dengan data nilai siswa kelas 5 semester dua tahun ajaran 2020-2021. Atribut yang dipilih untuk digunakan dalam penelitian ini yaitu nama, nilai pelajaran, uts, uas dan absen.

b) *Cleaning Data*

Pada tahap *cleaning* akan dilakukan proses membersihkan data yang terdapat *missing value* dan *duplikat* data yang dapat mempengaruhi hasil *clustering*. Menurut (Santosa, 2007) jika jumlah data yang akan digunakan besar maka *missing value* bisa dihapus dan jika jumlah data yang digunakan kecil atau terbatas, maka data *missing* bisa digantikan dengan nilai rata-rata dari atribut yang berkaitan.

c) **Data Transformasi**

Pada tahap ini dilakukan transformasi pada data nilai siswa kelas 5 tahun ajaran 2020-2021 dari tahap *cleaning* dimana atribut data mata pelajaran, uts, uas dan absen akan dirubah kedalam kode inisial yang unik agar mempermudah dalam implementasi kedalam perhitungan metode *k-means clustering*.

4. **Hitung Manual Algoritma K-Means**

Berikut ini akan dijelaskan tahapan-tahapan melakukan perhitungan dengan menggunakan metode *K-Means Clustering*.

1. Tentukan banyak nya *cluster* adalah 3 *cluster* yang akan dibuat. Data yang akan digunakan pada penelitian ini adalah data nilai siswa kelas 5 tahun ajaran 2020-2021.
2. Tentukan *centroid* tiap *cluster*. Dengan cara menentukan terlebih dahulu pusat *centroid* awal *cluster* yang diambil secara acak atau *random* dari data yang tersedia. Data pusat *centroid* awal *cluster* dapat dilihat pada Tabel 1 berikut:

Tabel 1. Data *Centroid* Awal

Data	iterasi 1	K01	K02	...	K13	K14
1	cluster 1	90	97	...	93	0
10	cluster 2	84	90	...	86	0
21	cluster 3	91	90	...	85	0

Penentuan *centorid* baru selanjutnya dapat menghitung nilai rata-rata pada tiap *cluster*. Jika *centroid* baru yang didapat berbeda dengan *centorid* pada iterasi sebelumnya maka proses iterasi akan dilanjutkan, sampai *centroid* yang baru akan sama dengan *centroid* di iterasi sebelumnya maka proses perhitungan *clustering* selesai.

3. Hitung jarak data dengan *centroid* menggunakan *Euclidean Distance*.

Setelah mendapatkan hasil jarak *centroid* 3 , Berikut perhitungan jarak data dengan ke 3 *centorid* dapat dilihat pada Tabel 2 berikut:

Tabel 2. Hasil jarak data dengan *centroid* awal

No.	Nama	<i>cluster 1</i>	<i>cluster 2</i>	<i>cluster 3</i>
1	siswa 1	0	807,959	792,567
2	siswa 2	1212,836	369,347	245,145
3	siswa 3	515,544	260,960	188,546
4	siswa 4	443,583	464,047	316,725
5	siswa 5	1098,723	312,080	341,108
6	siswa 6	529,593	397,020	183,821
7	siswa 7	319,050	366,907	281,294

8	siswa 8	1138,708	473,172	124,128
9	siswa 9	263,905	590,017	537,951
10	siswa 10	807,959	0	272,268
11	siswa 11	1615,112	617,421	624,605
12	siswa 12	621,258	206,650	329,763
13	siswa 13	2192,260	911,712	798,399
14	siswa 14	462,612	253,005	239,981
15	siswa 15	2016,296	573,064	490,664
16	siswa 16	153,169	868,058	754,617
17	siswa 17	3100,832	1323,447	1372,306
18	siswa 18	2003,117	578,932	527,689
19	siswa 19	3374,074	1374,877	1099,957
20	siswa 20	543,318	271,245	356,502
21	siswa 21	792,567	272,268	0

4. Kelompokkan data sesuai dengan *cluster* yaitu dengan menghitung jarak terdekat. Kelompok *cluster* pada *centroid* awal dapat dilihat pada Tabel 3 berikut:

Tabel 3. Hasil *cluster* pada data *centroid* awal

No.	Nama	C1	C2	C3
1	siswa 1	I		
2	siswa 2			I
3	siswa 3			I
4	siswa 4			I
5	siswa 5		I	
6	siswa 6			I
7	siswa 7			I
8	siswa 8			I
9	siswa 9	I		
10	siswa 10		I	
11	siswa 11		I	
12	siswa 12		I	
13	siswa 13			I
14	siswa 14			I
15	siswa 15			I
16	siswa 16	I		
17	siswa 17		I	
18	siswa 18			I
19	siswa 19			I
20	siswa 20		I	
21	siswa 21			I

5. Proses selanjutnya diulang ketahap nomor 2 dengan menggunakan *centroid* baru yang telah dihitung dari rata-rata tiap kelompok *cluster*.

Centroid baru dapat dilihat pada Tabel 4 berikut:

Tabel 4. *Centroid* baru iterasi kedua

iterasi 2	K01	K02	...	K13	K14
<i>cluster 1</i>	93	93	...	90	0
<i>cluster 2</i>	85	88	...	78	1
<i>cluster 3</i>	87	86	...	84	0

Langkah selanjutnya dengan menggunakan *centroid* baru untuk melakukan iterasi terus menerus sampai titik *centroid* sama dengan proses iterasi yang sebelumnya dan proses *clustering* berhenti. Proses iterasi pada data yang digunakan pada perhitungan *cluster* ini berhenti dan dinyatakan selesai pada iterasi ke 5.

5. Pengolahan Data Menggunakan RapidMiner

Sebelum melakukan pengolahan data, pada tahap ini dalam penggunaan algoritma *K-Means* pada *RapidMiner* terlebih dahulu tentukan *clusternya*. Pada data ini ditentukan dengan 3 *cluster* (k). Hasil dari proses *clustering* akan dihasilkan dalam bentuk *file* terpisah dalam tiap-tiap hasil *cluster*.

Pada *Description* maka akan menampilkan *cluster model*, *Cluster Model* yang diperoleh dari hasil pengujian terhadap data menggunakan *K-Means Clustering* terlihat pada Gambar 3 berikut:

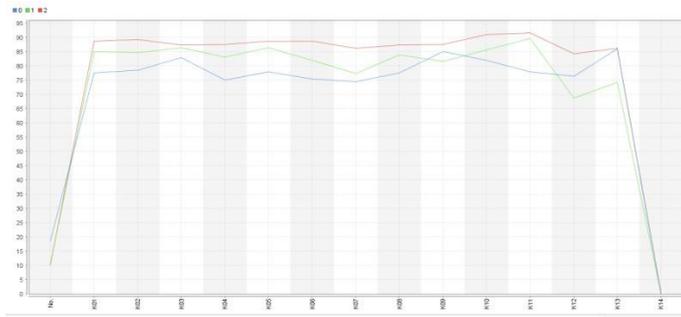
Cluster Model

```
Cluster 0: 2 items  
Cluster 1: 6 items  
Cluster 2: 13 items  
Total number of items: 21
```

Gambar 3. Hasil *Cluster Model*

Dapat dilihat pada *cluster model* yang dihasilkan terdiri dari *cluster_0*, *cluster_1* dan *cluster_2*. Pada *cluster_0* terdiri dari 2 *items*, *cluster_1* terdiri dari 6 *items* dan *cluster_2* terdiri dari 13 *items*. Dalam bentuk *root* memiliki 3 *cluster* yaitu *folder cluster_0*, *cluster_1* dan *cluster_2*. Untuk memudahkan melihat *items* dari masing-masing *cluster*, dapat dilihat pada Gambar 5.7 yang menampilkan *membership* dari masing-masing *cluster membership*.

Untuk melihat *cluster* mana yang paling tinggi, menengah dan rendah dapat dilihat pada Gambar 4 berikut:



Gambar 4. Tampilan diagram penggambaran *cluster*

Dari hasil tampilan diagram penggambaran *cluster* dapat ditentukan yang berwarna merah *cluster_2* menandakan *cluster* yang paling tinggi, yang berwarna hijau *cluster_1* menandakan menengah dan warna biru menandakan *cluster_0* yang rendah atau rendah.

KESIMPULAN

Dari penelitian yang telah dilakukan dapat disimpulkan bahwa algoritma *K-Means* untuk pemetaan siswa berprestasi di Sekolah Dasar Terang Bagi Bangsa Pati sebagai berikut:

1. Dengan menggunakan metode algoritma *K-Means Clustering*, dapat menentukan pemetaan prestasi siswa kelas 5 pada Sekolah Dasar Terang Bagi Bangsa Pati yaitu prestasi siswa tinggi, menengah dan rendah.
2. Dengan menginput data yang digunakan dalam penelitian ini yaitu berupa 15 atribut Nama, nilai pelajaran (ppkn, b.indo, ipa, ips, sbdp, mtk, b.jw, ssd, pjok, agama, b.ing), uts, uas dan absen. *Output* ditampilkan dari data yang digunakan yaitu berupa pemetaan data berdasarkan algoritma *K-Means*.
3. Hasil dari pengolahan menggunakan *Tools RapidMiner* dengan nilai data siswa kelas 5 merumuskan 3 *cluster*, *cluster_0* sebagai prestasi rendah yang terdapat 2 *items*, *cluster_1* sebagai prestasi menengah terdapat 6 *items* dan *cluster_2* sebagai prestasi tinggi yang terdapat 13 *items*.

DAFTAR PUSTAKA

- [1] N. Butarbutar, A. Perdana Windarto, D. Hartama, and Solikhun, "Komparasi Kinerja Algoritma Fuzzy C-Means dan K-Means dalam Pengelompokan Data Siswa Berdasarkan Prestasi Nilai Akademik Siswa (Studi Kasus : SMP Negeri 2 Pematangsiantar)," *J. Ris. Inf. Tek. Inform.*, vol. 1, no. 1, 2016.
- [2] F. Nur, M. Zarlis, and B. B. Nasution, "Penerapan Algoritma K-Means Pada Siswa Baru Sekolahmenengah Kejuruan Untuk Clustering Jurusan," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 1, no. 2, pp. 100–105, 2017, doi: 10.30743/infotekjar.v1i2.70.
- [3] R. A. Asroni, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Ilm. Semester Tek.*, vol. 18, no. 1, pp. 76–82, 2015.
- [4] H. Fonda, "Implementasi Text Mining Pada Twitter Dengan Algoritma K-Means Clustering Sebagai Dasar," vol. 9, no. 2, pp. 138–147, 2020.
- [5] Y. Prayoga, H. S. Tambunan, and I. Parlina, "Penerapan Clustering Pada Laju Inflasi Kota Di Indonesia," *J. penerapan kecerdasan buatan*, vol. 1, no. 1, pp. 24–30, 2019.

[1]