

ADAPTIVE AFFINITY PROPAGATION UNTUK PENGELOMPOKAN KEHADIRAN MAHASISWA PEMBELAJARAN DARING SELAMA PANDEMI DENGAN NILAI AKHIR

Millati Izzatillah¹⁾, Fauzan Natsir²⁾, Siti Anisah³⁾ dan Opitasari⁴⁾

^{1, 2,3,4)} Prodi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Indraprasta PGRI
e-mail: mizzatillah@gmail.com¹⁾, fauzan.natsir@gmail.com²⁾, anis.siti.ssa@gmail.com³⁾, opitasari@gmail.com⁴⁾

ABSTRAK

Clustering merupakan metode pengelompokan data dengan mudah dan bertujuan untuk melihat korelasi antar data atribut. Clustering adalah proses pengelompokan titik-titik data yang diambil berdasarkan kesamaan nilai untuk menentukan pusat cluster (exemplar). Affinity Propagation (AP) dan Adaptive Affinity Propagation (AAP) adalah algoritma clustering yang dapat menghasilkan jumlah cluster, anggota cluster, dan contoh dari setiap cluster. Penelitian ini dilakukan untuk mengetahui algoritma yang lebih efektif dalam pengelompokan data. AAP merupakan pengembangan dari algoritma AP yang telah melakukan beberapa perbaikan dalam clustering data. Peneliti juga mengimplementasikan dan menguji kedua algoritma tersebut dengan MATLAB 2021 menggunakan data presensi siswa dalam pembelajaran daring di masa pandemi dengan data nilai akhir sebanyak 500 data. Hasil dari penelitian ini adalah algoritma AP menghasilkan lebih banyak klaster daripada algoritma AAP, klaster AP 5 menggunakan preferensi minimum dan 8 klaster menggunakan preferensi median, klaster AAP 3 menggunakan preferensi minimum dan 2 klaster menggunakan preferensi median. Pemrosesan data runtime AAP lebih cepat daripada AP menggunakan preferensi minimum atau median. Algoritma terbaik dalam clustering data secara optimal adalah Adaptive AP. Hasil lainnya adalah terdapat korelasi antara pengelompokan kehadiran siswa terhadap pembelajaran daring di masa pandemi dengan nilai akhir.

Kata Kunci: Algoritma Clustering, Pengelompokan Data, Propagasi Afinitas, Propagasi Afinitas Adaptif, Matlab.

ABSTRACT

Clustering is a method of clustering data easily and has the aim of seeing the correlation between attribute data. Clustering is the process of grouping data points that are taken based on the similarity value to determine the center of the cluster (exemplar). Affinity Propagation (AP) and Adaptive Affinity Propagation (AAP) are clustering algorithms that can generate the number of clusters, cluster members, and exemplar of each cluster. This research is made to find out the more effective algorithm in clustering data. AAP is the development of AP algorithm that has made some improvements in clustering data. Researchers also implemented and tested both algorithm with MATLAB 2021 using student data of attendance in online learning during a pandemic with a final grade of 500 data. The result of this research was AP algorithm produced more clusters than AAP algorithm, AP 5 clusters using minimum preference and 8 clusters using median preference, AAP 3 clusters using minimum preference and 2 clusters using median preference. Runtime AAP data processing is faster than AP using either minimum or median preference. The best algorithm in clustering data optimally is Adaptive AP. Another result is there is correlation between grouping student attendance in online learning during the pandemic with final grade.

Keywords: Clustering Algorithm, Data Grouping, Affinity Propagation, Adaptive Affinity Propagation, Matlab.

I. PENDAHULUAN

Clustering merupakan salah satu teknik dari salah satu fungsi data mining, *algoritma clustering* adalah algoritma untuk mengelompokkan sejumlah data ke dalam kelompok data (*cluster*) tertentu [1]. *Cluster* adalah sekelompok atau kumpulan objek data yang mirip satu sama lain dalam *cluster* yang sama dan berbeda dengan objek *cluster* yang berbeda [2].

Dengan kata lain *cluster* adalah teknik pengelompokan data yang memiliki tingkat kemiripan tertentu sehingga dapat berada dalam satu kelompok yang sama.

Affinity Propagation (AP) adalah algoritma yang mengambil sebagai *input* ukuran kesamaan antara pasangan titik data, pesan nilai nyata dipertukarkan antara titik data hingga satu set salinan berkualitas tinggi dan *cluster* yang sesuai secara bertahap muncul,

keuntungannya adalah menemukan *cluster* dengan banyak kesalahan lebih rendah daripada metode lain dan dilakukan dengan waktu yang relatif singkat [3].

Seiring dengan perkembangannya, *Affinity Propagation* (AP) tidak selalu menghasilkan *cluster* yang ideal. *Affinity Propagation* (AP) memiliki dua keterbatasan: (1) AP sulit mengetahui parameter “*preference*” sehingga sulit menghasilkan solusi *clustering* yang optimal. (2) Osilasi tidak dapat dihapus secara otomatis jika terjadi [4].

Keterbatasan AP dapat menghasilkan *cluster* yang belum optimal. Maka untuk mengatasi keterbatasan tersebut, *Affinity propagation* (AP) dikembangkan menjadi *Adaptive Affinity propagation* (AAP) [5].

AP adaptif dapat menghilangkan osilasi yang akan terjadi dengan menghilangkannya (*adaptive damping*), jika *adaptive damping* gagal maka *Adaptive AP* akan menghindari osilasi (disebut *adaptive escaping*) dan setelah itu akan dilakukan *adaptive search* untuk menemukan solusi *clustering* yang paling optimal (disebut *adaptive scanning*) [4].

Penelitian ini menguji dan mengimplementasikan algoritma AP dan AAP dengan membandingkan hasilnya dan untuk mengetahui seberapa efektif algoritma tersebut dalam pengelompokan data. Pengujian dilakukan dengan menggunakan data kehadiran siswa dalam pembelajaran daring di masa pandemi dengan nilai akhir sebanyak 500 siswa. Implementasi dilakukan dengan *coding* pada perangkat lunak pengolah data MATLAB 2021.

II. LANDASAN TEORI

A. *Affinity Propagation* (AP)

Pengelompokan berbasis contoh, seperti metode pengelompokan -centers dan -median yang populer, mempartisi dataset dengan mengidentifikasi subset elemen representatif sebagai contohnya sehingga jumlah jarak antara titik data dan contoh dapat diminimalkan [6].

Metode analisis pengelompokan tradisional biasanya dimulai dengan langkah inisialisasi di mana algoritma memilih pusat data awal sebagai contoh dan mengalokasikan titik data lain berdasarkan jarak ke contoh. Jelas bahwa pemilihan awal yang berbeda menghasilkan hasil pengelompokan yang berbeda. Sebaliknya, AP berjalan berdasarkan mekanisme yang sama sekali berbeda. Pertama, semua titik data dianggap sebagai contoh potensial dan dipandang sebagai simpul dalam jaringan. Kedua, sejumlah pesan bernilai nyata ditransmisikan secara iteratif di sepanjang tepi jaringan sehingga satu set contoh yang relevan dan kluster yang sesuai dapat diidentifikasi [7].

AP mengambil sebagai input matriks kesamaan bernilai nyata antara titik data. Misalkan $\{s_{ij}\}$, $i = 1, \dots, N$, $j = 1, N$, adalah himpunan N^2 variabel bernilai

riil dimana s_{ij} menunjukkan kesamaan antara dua objek x_i dan x_j di dalamnya. AP mendefinisikan s_{ij} sebagai negatif dari kuadrat jarak Euclidean mereka; yaitu, $s(i, j) = -\|x_i - x_j\|_2^2$, $i = j$. Kemiripan diri s_{kk} disebut sebagai “preferensi” yang memengaruhi probabilitas satu titik menjadi contoh. Jika tidak ada pengetahuan apriori, preferensi ditetapkan ke nilai umum sehingga setiap titik data dianggap sebagai contoh potensial dengan probabilitas yang sama. Preferensi akan mempengaruhi hasil total *cluster*. Preferensi dibagi menjadi dua jenis, yaitu median atau minimum [8].

$$P = med, \text{ or } P = min; \quad (1)$$

Poin data yang akan dilewatkan oleh dua nilai ini, dengan nilai responsibility $s(i, j)$ dan availability $r(i, j)$. Kesamaan hasil yang dimaksimalkan dapat membuat kluster secara optimal. *Responsibility* menunjukkan seberapa cocok data j sebagai contoh untuk data i . Akumulasi bukti tercermin pada data i yang sesuai untuk memilih data j sebagai contoh yaitu *availability* $a(i, j)$.

$$r(i, j) = \begin{cases} s(i, j) - \max_{k \neq i} \{a(i, k) + s(i, k)\} & (i \neq j) \\ s(i, j) - \max_{k \neq i} \{s(i, k)\} & (i = j) \end{cases} \quad (2)$$

$$a(i, j) = \begin{cases} \min\{0, r(i, j)\} + \sum_{k \neq i, j} \max\{0, r(k, j)\} & (i \neq j) \\ \sum_{k \neq i, j} \max\{0, r(k, j)\} & (i = j) \end{cases} \quad (3)$$

Namun, AP juga memenuhi beberapa batasan. Nilai spesifik dari “preferensi”, untuk prosedur pengelompokan, adalah pedang bermata dua. Frey dan Dueck [20] menyarankan pengaturan nilai preferensi bersama sebagai median kesamaan input yang menghasilkan jumlah cluster yang moderat. Sulit untuk menentukan nilai preferensi yang sesuai di mana nilai yang berbeda mengarah pada hasil yang sama sekali berbeda jika tidak ada pengetahuan apriori. Selain itu, faktor redaman memperoleh nilai yang sesuai.

III. METODOLOGI PENELITIAN

A. Tahapan Penelitian

Dalam proses pelaksanaan penelitian, tahapan penelitian dibagi menjadi 4 tahapan penelitian dalam penelitian implementasi yang terdiri dari:

1. Studi Pendahuluan

Langkah awal penelitian ini adalah mempelajari masalah yang ada di sekolah kemudian menentukan ruang lingkup masalah, latar belakang masalah, dan mempelajari terkait masalah serta cara mencari solusi dari masalah yang terjadi.

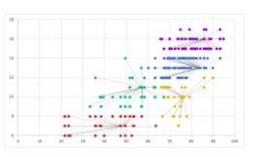
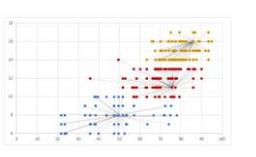
Studi Pendahuluan juga melakukan studi pustaka yang digunakan dalam penulisan penelitian ini yaitu mengumpulkan bahan referensi mengenai metode klasifikasi dari berbagai buku, jurnal, dan beberapa referensi lainnya.

2. Pengumpulan data

Melakukan observasi di sekolah untuk mengetahui

informasi yang dibutuhkan, penulis mengumpulkan data. Jumlah data yang digunakan untuk clustering adalah 500 data, data tersebut berdasarkan nilai rata-rata dan juga kehadiran siswa kelas 7 hingga kelas 9 selama 1 semester sejak pandemi. data diambil secara acak, digunakan sebagai variabel penentu antara lain 2 variabel yaitu nilai akhir dan kehadiran.

TABLE I. RECORDS FINAL GRADE AND ATTENDANCE

Final Grade	Attendance
69,16	12
47,5	8
70	12
85,83	15
70	14
	
50000	50000
5	3
4,02 second	1,05 second
50, 100, 298, 441, 480	69, 190, 472

3. Pemrosesan Data dengan Data Mining

Pada tahap pengolahan data, terlebih dahulu mengidentifikasi permasalahan yang ada. Tahap selanjutnya adalah analisis masalah dengan tujuan agar penulis mengetahui dan memperoleh gambaran yang jelas bagaimana bentuk penyelesaiannya dan algoritma apa yang dapat digunakan untuk penyelesaiannya.

4. Analisis Hasil

Penulis menganalisis hasilnya menggunakan pembuatan kode pada *software* aplikasi MATLAB 2021, sehingga algoritma *Affinity Propagation* dan *Adaptive Affinity Propagation* dapat diuji dengan data untuk menghasilkan *cluster*.

5. Kesimpulan

Tahap ini merupakan kesimpulan dari penelitian yang dilakukan dalam mengklasifikasikan.

B. Instrumen Penelitian

Berdasarkan jenis data yang diperoleh penulis dalam penelitian ini, maka penulis menggunakan metode kuantitatif dengan menggunakan metode penelitian deskriptif untuk menggambarkan hasil penelitian yang dilakukan. Metodologi penelitian kuantitatif yang menggunakan data numerik dan menekankan proses penelitian pada pengukuran hasil yang objektif dengan menggunakan analisis statistik Gambar yang terdiri dari lebih dari satu gambar ditampilkan berdampingan,

atau ditumpuk.

Pada dasarnya, metode analisis data meliputi:

1. Analisis masalah. Berkaitan dengan mengidentifikasi kebutuhan dalam sebuah penelitian. Akan dijelaskan permasalahan yang terjadi dan bagaimana tahapan penyelesaian permasalahan tersebut nantinya.
2. Metode Analisis Algoritma Affinity Propagation Dan Adaptive Affinity Propagation. Bagian ini membandingkan kinerja pengelompokan antara AP adaptif (AAP) dan algoritma AP (AP). Item kinerja pengelompokan meliputi: apakah AAP dapat menghilangkan osilasi (jika terjadi osilasi) secara otomatis sehingga memberikan hasil pengelompokan yang benar, apakah AAP dapat memberikan hasil pengelompokan yang benar berdasarkan indeks Silhouette (atau teknik validasi kluster). AAP dan AP menggunakan lam=0,5 awal yang sama (tetapi lam=0,8 dalam percobaan rute perjalanan), dan AP menggunakan p=pm tetap dan maks=5000. Biarkan kumpulan data menjadi matriks n×d X={xi}, di mana xi adalah d-dimensi. berdasarkan jarak Euclidean, sedangkan untuk data ekspresi gen digunakan koefisien Pearson sebagai ukuran kesamaan, yaitu hubungan linier antara dua sampel xi dan xj[4].

$$r(i,j) = \frac{\sum_{l=1}^d (X_{il} - \bar{X}_i)(X_{jl} - \bar{X}_j)}{\sqrt{\sum_{l=1}^d (X_{il} - \bar{X}_i)^2} \sqrt{\sum_{l=1}^d (X_{jl} - \bar{X}_j)^2}}$$

Propagasi afinitas bekerja melalui beberapa tahap. Tahapan-tahapan tersebut dijelaskan dengan flowchart berikut.

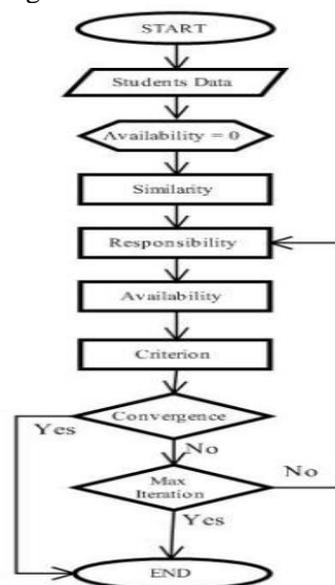


Fig. 1. Program Flowchart Affinity Propagation Algorithm

AAP merupakan pengembangan dari algoritma AP yang mengusulkan perbaikan untuk mencapai hasil clustering yang lebih optimal dan mengatasi keterbatasan AP. Oleh karena itu, beberapa tahap awal AP dan AAP sama dengan AP (1) input, (2) kesamaan, (3) tanggung jawab, (4) ketersediaan dan (5) kriteria. Perbanyak Adaptive Affinity (Adaptive AP) bekerja melalui beberapa tahapan yang dijelaskan oleh flowchart program pada Gambar 2.

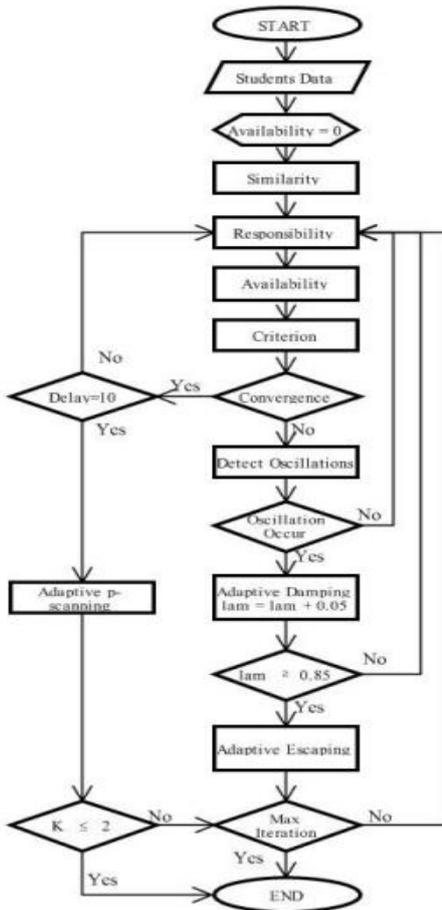


Fig. 2. Program Flowchart Adaptive Affinity Propagation Algorithm

Setelah didapatkan jumlah cluster dari setiap iterasi, tahap selanjutnya adalah menentukan apakah jumlah cluster sudah optimal. Jumlah cluster yang optimal akan terjadi ketika mencapai nilai konvergen. Nilai konvergensi akan dicapai dengan ketentuan sebagai berikut: Jumlah cluster akan selalu sama setelah iterasi ke-50 (50 iterasi), (ukuran window = 40, ditambah delay = 10). Contoh (pusat cluster) tidak berubah setelah banyak iterasi.

Pada AAP, setelah dilakukan pengecekan konvergensi, terdapat beberapa tahapan sebelum melakukan pengecekan iterasi maksimal (maxit = 50000 maksimal iterasi [8]). Jika k mencapai konvergensi dengan iterasi tambahan dy (delay = 10) akan melakukan p-scanning adaptif. Jika tidak konvergen dengan iterasi tambahan (delay ≠ 10) maka akan dilanjutkan iterasi tanggung jawab. Jika tidak juga mencapai konvergen makan

akan memeriksa adanya isolasi. Mendeteksi osilasi terlalu rumit untuk dijelaskan. Jadi, deteksi osilasi dengan mendefinisikan fitur non-osilasi. Kondisi non-osilasi ketika contoh berkurang atau tidak berubah selama proses iterasi. Contoh yang mengalami penurunan atau perubahan akan terjadi konvergensi[8]. Osilasi terjadi ketika jumlah salinan yang dihasilkan bertambah dan mengubah 2/3 ukuran jendela, sehingga lambda akan bertambah. Nilai awal lambda adalah 0,5. Teknik ini akan bekerja ketika lambda telah mencapai 0,85 atau lebih tinggi. Lambda besar yang membawa efek kecil menunjukkan bahwa osilasi terjadi di bawah p yang diberikan, jadi alternatifnya adalah mengurangi p dari p yang diberikan untuk menghindari osilasi. AAP Escaping dirancang untuk menghindari osilasi saat AAP Damping gagal menekan osilasi.

IV. HASIL DAN PEMBAHASAN

Implementasi algoritma AP dan AAP ini menggunakan aplikasi pengkodean online MATLAB. Data siswa yang digunakan dalam penelitian ini adalah 500 data. Terdapat kehadiran siswa pada pembelajaran daring di masa pandemi dengan nilai akhir. Algoritma AP dan AAP diuji beberapa kali dengan dua jenis preferensi (Minimum dan Median). Untuk setiap kali mendapatkan hasil yang berbeda.

TABLE II. TESTING RESULT OF AP AND AAP ALGORITHM WITH MINIMUM PREFERENCE

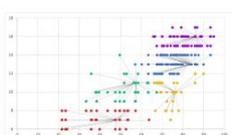
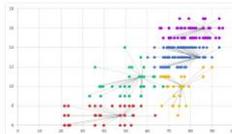
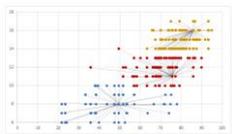
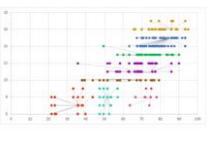
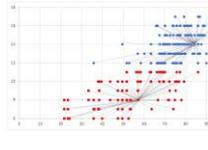
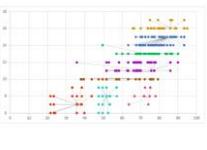
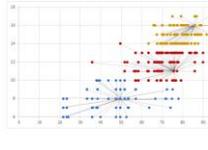
Table Head	Testing Result With Minimum Preference	
	Affinity Propagation (AP)	Adaptive Affinity Propagation (AAP)
First Testing (Plotting)		
Converge Iteration	8340	1123
Number of Cluster	5	3
Runtime	3,75 second	0,98 second
Exemplar	50, 91, 298, 441, 473	69, 191, 473
Second Testing (Plotting)		
Maximum Iteration	50000	50000
Number of Cluster	5	3
Runtime	4,02 second	1,05 second
Exemplar	50, 100, 298, 441, 480	69, 190, 472

TABLE III. TESTING RESULT OF AP AND AAP ALGORITHM WITH MEDIAN PREFERENCE

Table Head	Testing Result With Median Preference	
	Affinity Propagation (AP)	Adaptive Affinity Propagation (AAP)
First Testing (Plotting)		
Converge Iteration	909	2301
Number of Cluster	8	2
Runtime	1,75 second	2,98 second
Exemplar	25, 73, 139, 326, 332, 330, 464, 482	100, 475
Second Testing (Plotting)		
Maximum Iteration	50000	50000
Number of Cluster	8	2
Runtime	4,02 second	2,13 second
Exemplar	20, 57, 90, 150, 270, 300, 420, 490	150, 450

Menguji algoritma ini untuk setiap preferensi adalah dua kali. Pengujian pertama jumlah iterasi sampai tercapai konvergensi. Jumlah pengujian kedua iterasi hingga maksimum (maxit=50000 iterasi). Hasil pengujian menggunakan preferensi minimum menghasilkan jumlah kluster yang sama, AP sebanyak 5 kluster dan AAP sebanyak 3 kluster. Runtime plotting data AAP 0,98 detik lebih cepat dibandingkan AP 3,75 detik. Hasil pengujian menggunakan preferensi median juga menghasilkan jumlah cluster AP yang sama yaitu 8 dan AAP adalah 2. Runtime plotting data AAP juga lebih cepat dari AP pada preferensi median. Jumlah cluster yang dihasilkan oleh algoritma AP lebih besar dari algoritma AAP. Dari tampilan data hasil *clustering* tersebut, data kehadiran siswa dan nilai akhir memiliki korelasi. Jumlah kehadiran kelas *online* selama pandemi mempengaruhi nilai akhir yang diperoleh siswa.

V. KESIMPULAN

Algoritma AP menghasilkan kluster lebih banyak dari pada algoritma AAP, kluster AP 5 menggunakan preferensi minimum dan 8 kluster menggunakan preferensi median, kluster AAP 3 menggunakan preferensi minimum dan 2 kluster menggunakan preferensi median. *Runtime* memplot data AAP juga lebih cepat dari AP dalam preferensi median atau minimum.

Data yang digunakan memiliki hubungan yaitu data

nilai rata-rata mahasiswa dengan rekapitulasi kehadiran siswa selama satu semester. Nilai rata-rata diperoleh dari beberapa nilai mata pelajaran yang dihitung menurut jumlah siswa pada setiap kelas. Jumlah kehadiran kelas *online* selama pandemi mempengaruhi nilai akhir yang diperoleh siswa.

Hasil akhir AAP menghasilkan jumlah *cluster* yang optimal dengan proses *runtime* tercepat saat melakukan plotting data. Dengan kata lain keberhasilan AAP meningkatkan AP yang memiliki beberapa keterbatasan.

DAFTAR PUSTAKA

- [1] M. L. Sibuea and A. Safta, "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering," *Jurteks*, vol. 4, no. 1, pp. 85–92, 2017, doi: 10.33330/jurteks.v4i1.28.
- [2] B. M. Metisen and H. L. Sari, "Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.
- [3] B. K. Amijaya, M. T. Furqon, and C. Dewi, "Clustering Titik Panas Bumi Menggunakan Algoritme Affinity Propagation," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 10, pp. 3835–3842, 2018.
- [4] K. J. Wang, J. Y. Zhang, D. Li, X. N. Zhang, and T. Guo, "Adaptive affinity propagation clustering," *Zidonghua Xuebao/Acta Autom. Sin.*, vol. 33, no. 12, pp. 1242–1246, 2007, doi: 10.1360/aas-007-1242.
- [5] J. O. Ullman, *Mining of Massive Datasets*. 2014.
- [6] I. E. Givoni and B. J. Frey, "A binary variable model for affinity propagation," *Neural Comput.*, vol. 21, no. 6, pp. 1589–1600, 2009, doi: 10.1162/neco.2009.05-08-785.
- [7] I. E. Givoni, C. Chung, and B. J. Frey, "Hierarchical Affinity Propagation.pdf."
- [8] R. Refianti, A. B. Mutiara, A. Juarna, and A. Suhendra, "A preference model on adaptive affinity propagation," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 3, pp. 1805–1813, 2018, doi: 10.11591/ijece.v8i3.pp1805-1813.
- [9] Natsir, F., & Sihombing, R. A. (2022). Penerapan Metode User Centered Design pada Rancangan User Interface Marketplace Pemasaran Produk Olahan Perikanan. *Journal of Practical Computer Science*, 2(2), 56-63.