

**PENGGUNAAN *K-MEANS METHOD* DALAM KLASIFIKASI TINGKAT
KESEJAHTERAAN PENDUDUK (STUDI KASUS: KANTOR KEPALA DESA
PONDOK BUNGUR)**

Sylvia Anggraini¹, Akmal Nasution^{2*}, Mustika Fitri Larasati²

¹Mahasiswa Prodi Sistem Informasi, STMIK Royal

²Dosen Prodi Sistem Informasi, STMIK Royal

*email: nst.akmal@gmail.com

Abstract: The quality of family life reflects community welfare, measured through various aspects of life such as education, employment, income, ownership, living conditions, and access to clean water and electricity. To determine the level of welfare in the community, three criteria from all these variables are used to differentiate whether they belong to a high, medium, or low cluster. Pondok Bungur, a village in Asahan Regency with 10 hamlets, where most residents have low incomes, lack secondary necessities, use non-PDAM water, and have low electricity access. Due to the lack of clear classification regarding welfare, social aid often misses the target. As a solution, data mining techniques with the k-means algorithm are applied to classify the level of welfare in the village. This study found that 2 hamlets are at a high welfare level (20%), 3 hamlets at a medium level (30%), and 5 hamlets at a low level (50%). Hamlets 2 and 5 are the most prosperous, while Hamlets 1, 4, and 9 have a medium level of welfare. The remaining 50% of the hamlets that are still at a low level need to be prioritized for assistance.

Keywords: *Data Mining, K-Means, Classification, Population Welfare*

Abstrak: Kualitas hidup keluarga mencerminkan kesejahteraan masyarakat, yang diukur melalui berbagai aspek kehidupan seperti pendidikan, pekerjaan, penghasilan, kepemilikan, kondisi tempat tinggal, serta akses ke air bersih dan listrik. Untuk menentukan tingkat kesejahteraan di masyarakat, tiga kriteria dari semua variabel tersebut digunakan untuk membedakan apakah mereka berada pada *cluster* tinggi, menengah, atau rendah. Pondok Bungur, sebuah desa di Kabupaten Asahan dengan 10 dusun, kebanyakan penduduknya memiliki pendapatan rendah, tidak memiliki barang kebutuhan sekunder, menggunakan air bukan dari PDAM, dan memiliki akses daya listrik rendah. Karena belum ada klasifikasi yang jelas mengenai kesejahteraan, bantuan sosial sering kali tidak tepat sasaran. Sebagai solusi, teknik data mining dengan algoritma *k-means* diterapkan untuk mengklasifikasikan tingkat kesejahteraan di desa tersebut. Penelitian ini menemukan bahwa 2 dusun berada pada tingkat kesejahteraan tinggi (20%), 3 dusun pada tingkat menengah (30%), dan 5 dusun pada tingkat rendah (50%). Dusun 2 dan 5 merupakan yang paling sejahtera, sementara Dusun 1, 4, dan 9 memiliki tingkat kesejahteraan menengah. Sisa 50% dusun yang masih berada pada tingkat rendah perlu diprioritaskan dalam pemberian bantuan.

Kata kunci: *Data Mining, K-Means, Klasifikasi, Kesejahteraan Penduduk*

PENDAHULUAN

Dalam era globalisasi dan perkembangan teknologi, penilaian tingkat kesejahteraan penduduk menjadi suatu aspek penting dalam upaya pembangunan suatu daerah. Peningkatan kesejahteraan masyarakat bukan hanya sekedar dari segi ekonomi, melainkan juga mencakup faktor-faktor sosial, kesehatan, dan lingkungan. Penduduk merupakan unit masyarakat yang memegang peran strategis dalam menciptakan keseimbangan antara pertumbuhan ekonomi, kesejahteraan sosial, dan kelestarian lingkungan[1]. Permasalahan kesejahteraan ini memang tidak akan sepenuhnya dapat diatasi namun hal ini harus bisa ditekan serendah mungkin dan ditangani dengan serius untuk mewujudkan pembangunan yang merata ke segala lapisan masyarakat. Tingkat kesejahteraan dapat diukur berdasarkan pendapatan perkapita suatu negara/daerah tertentu. Indonesia sendiri yang memiliki jumlah penduduk 280,73 juta jiwa pada tahun 2023 berdasarkan data dari Dukcapil. Untuk mengukur pendapatan per kapita Indonesia kita membutuhkan Total Pendapatan Nasional yang dihitung dari Produk Nasional Bruto (PNB) atau Produk Domestik Bruto (PDB).

Berdasarkan data dari BPS diketahui bahwa PDB per kapita Indonesia untuk tahun 2023 adalah Rp. 74.964.701. Secara global, PDB per kapita tersebut setara dengan sekitar USD 4919,7 (berdasarkan kurs saat ini). Menurut klasifikasi Bank Dunia, negara-negara dibagi menjadi beberapa kategori berdasarkan pendapatan nasional bruto (PNB) per kapita, yaitu pendapatan tinggi (lebih dari USD 12,536 per tahun), pendapatan menengah-atas (USD 4,046 hingga USD 12,535 per tahun), pendapatan menengah-bawah (USD 1,036 hingga USD 4,045 per tahun), dan pendapatan rendah (Kurang dari USD 1,035 per tahun). Dengan pendapatan per kapita sekitar USD 4.919,7 per tahun, Indonesia berada di bawah ambang batas kategori negara berpendapatan menengah-atas menurut standar ini. Meskipun secara global Indonesia masuk dalam pendapatan menengah-atas, namun untuk mengukur kesejahteraan tentu dapat dinilai juga dari kondisi lainnya dan biaya hidup yang berbeda disetiap daerah. Level kesejahteraan masyarakat menunjukkan bagaimana kemampuan masyarakat di berbagai aspek kehidupan. Level ini dapat diukur melalui berbagai indikator seperti tingkat pendidikan, jenis pekerjaan, jumlah penghasilan, kepemilikan aset, kondisi perumahan, serta akses terhadap air bersih dan penerangan yang layak [2].

Pondok bungur merupakan salah satu desa yang ada di Kecamatan Rawang Panca Arga, Kabupaten Asahan. Berdasarkan data pada BPS Sumut, Asahan sendiri memiliki jumlah penduduk 802,563 jiwa pada tahun 2023. Sementara Desa Pondok Bungur di tahun 2023 memiliki jumlah penduduk 5013 jiwa yang terdiri dari 10 dusun, dimana mayoritas penduduknya berpendidikan rendah, memiliki penghasilan yang tidak banyak, menganggur, dan kekurangan barang kebutuhan sekunder. Mereka juga menggunakan air yang bukan dari PDAM serta memiliki akses yang terbatas terhadap listrik. Rata-rata pendapatan perbulan desa Pondok Bungur adalah Rp. 1,88 juta per bulan. Sehingga jika dihitung pendapatan per kapita daerah Desa Pondok Bungur hanya sekitar Rp. 4.500,29 per tahun. Jika di bandingkan data global nilai ini tentunya sangat rendah, hal ini dapat mempengaruhi kesejahteraan penduduk untuk kedepannya. Meskipun jika dibandingkan secara global tingkat kesejahteraan penduduk di desa Pondok Bungur termasuk rendah, diperlukan peran yang lebih besar dari pemerintah desa untuk melaksanakan upaya dan tindakan-tindakan yang dapat meningkatkan

kesejahteraan masyarakat desa tersebut. Salah satunya dengan dilakukannya klasifikasi terhadap penduduk desa Pondok Bungur agar pemerintah setempat dapat memetakan dusun mana dengan penduduk yang tingkat kesejahteraan yang paling rendah sebagai prioritas dalam pemberian bantuan sosial.

Satu isu yang dihadapi oleh pemerintah desa Pondok Bungur saat ini adalah kesulitan dalam mengidentifikasi warga berdasarkan tingkat kesejahteraannya. Oleh karena itu, pemerintah desa perlu melakukan klasifikasi penduduk berdasarkan kriteria yang telah ditentukan untuk mendapatkan hasil yang relevan. Untuk mendapatkan hasil yang relevan diperlukan metode yang dapat menggali informasi dari data tersebut, metode yang dapat digunakan yaitu metode Data Mining. Data mining merupakan metode yang digunakan dalam pengolahan data berskala besar oleh karena itu data mining memiliki peranan yang sangat penting dalam beberapa bidang kehidupan diantaranya yaitu bidang industri, bidang keuangan, cuaca, ilmu dan teknologi [3]. *Data mining* memiliki beberapa metode, salah satu di antaranya adalah *K-Means clustering*. *K-Means* adalah algoritma pengklusteran yang mengelompokkan data berdasarkan pada nilai *centroid* (titik pusat cluster). *K-Means* menggunakan fungsi jarak pada pengklusteran data terhadap *centroid* [5]. Melalui metode *K-Means* nanti klasifikasi dapat dibagi menjadi 3 bagian yaitu tinggi, menengah, dan rendah dimana klasifikasi rendah akan mendapatkan prioritas bantuan sosial terlebih dahulu. Untuk itu penelitian ini dilakukan dengan tujuan untuk mempelajari dan mengetahui tingkat kesejahteraan penduduk di desa pondok bungur, untuk mengetahui bagaimana mengelola data penduduk yang banyak menggunakan metode *K-means*, dan untuk membuat aplikasi data mining berbasis *web* yang menggunakan metode *K-means* dalam mengklasifikasi kesejahteraan penduduk di Desa Pondok Bungur.

METODE

Metode *K-Means* merupakan algoritma yang menggunakan pendekatan non-hirarki atau partisi. Tujuan algoritma ini adalah melakukan partisi data menjadi beberapa kelompok atau *cluster* dimana data yang memiliki tingkat kemiripan karakteristik yang tinggi ditempatkan pada satu *cluster* yang sama. Metode bekerja dengan memanfaatkan masukan berupa nilai *k* yang merupakan representasi dari jumlah *cluster* yang harus *K-Means* buat. Nilai *k* juga akan menentukan berapa banyak titik *centroid* yang akan digunakan untuk membentuk *cluster* selama proses *clustering* nanti [4]. *K-Means* adalah algoritma pengklusteran yang mengelompokkan data berdasarkan pada nilai *centroid* (titik pusat cluster). *K-Means* menggunakan fungsi jarak pada pengklusteran data terhadap *centroid* [5].

Dalam melakukan *clustering* langkah-langkah dengan metode *K-Means* adalah sebagai berikut [6] :

1. Tentukan jumlah kluster *K*.
2. Inisialisasi pusat kluster dengan memberikan nilai acak pada *K* pusat kluster, yang dapat dilakukan dengan beberapa cara, salah satunya adalah secara acak.
3. Alokasikan setiap data atau objek ke kluster yang paling dekat. Kedekatan antara dua objek atau antara data dan pusat kluster diukur berdasarkan jarak. Jarak tiap data ke pusat kluster dihitung pada tahap ini, dengan menggunakan formula

Euclidean Distance untuk menentukan keanggotaan data pada kluster. Rumus formulanya seperti terlihat pada rumus (1):

$$D(i, j) = \sqrt{(X1i - Y1i)^2 + (X2i - Y2i)^2 + \dots + (Xki - Ykj)^2} \quad (1)$$

Penjelasan :

D(i, j) : Jarak dari data ke i ke pusat *cluster* j

Xki : Data ke i pada atribut data ke k

Ykj : Titik pusat ke j pada atribut ke k

4. Hitung ulang pusat kluster berdasarkan anggota kluster saat ini. Pusat kluster dihitung sebagai rata-rata dari semua data atau objek dalam kluster tersebut, atau menggunakan median dari data kluster.
5. Ulangi proses penugasan data ke pusat kluster yang baru. Jika pusat kluster stabil dan tidak mengalami perubahan, maka proses pengklusteran dianggap selesai. Jika tidak, kembali ke langkah 3 dan ulangi sampai pusat kluster stabil.

Setelah proses pengklusteran untuk masing-masing nilai K selesai, untuk menentukan jumlah kluster terbaik bisa menggunakan *Davies-Bouldin Index* (DBI). Metode ini dirancang untuk mengoptimalkan jarak antarkluster, sekaligus meminimalkan jarak antarobjek dalam kluster yang sama. Kluster yang paling ideal adalah yang memiliki nilai DBI paling rendah, menunjukkan kualitas pengklusteran yang optimal [7]. Nilai DBI dirumuskan pada rumus (2), rumus (3), dan rumus (4) berikut:

$$DBI = \frac{1}{k} \sum_{a=1}^k R_a \quad (2)$$

dengan,

$$R_a = \max_{b=1, \dots, k, a \neq b} R_{ab}, R_{ab} = \frac{S_a + S_b}{d(V_a, V_b)} \quad (3)$$

dimana:

k = Jumlah *cluster*

R_{ab} = Ukuran kemiripan antara *cluster* ke-a dan *cluster* ke-b

S_a = Ukuran dispersi *cluster* ke-a, a = 1, 2, ..., k

$$S_a = \left[\frac{1}{n_a} \sum_{T_i \in C_{a,i=1}}^{n_a} d^2(T_i, V_a) \right]^{1/2}, d^2(T_i, V_a) = (d(T_i, V_a))^2 \quad (4)$$

dimana:

n_a = banyaknya anggota *cluster* ke-a,

a = 1, 2, ..., k C_a = *Cluster* ke-a

T_i = Anggota ke-i pada *cluster* ke-a, a = 1, 2, ..., k

HASIL DAN PEMBAHASAN

Data awal yang dijadikan sebagai bahan masukan dalam penelitian ini berhubungan dengan data penduduk desa pondok bungur seperti terlihat pada tabel 1.

Tabel 1. Data Awal Penduduk Desa Pondok Bungur 2023

Dusun	Penduduk		Pekerjaan	Rumah		
	Jumlah Penduduk	Jumlah Kepala Keluarga	Rata-rata Penghasilan (per bulan) (jt)	Rata-rata Daya Listrik	Jumlah Pengguna Air PAM	% Pengguna PAM
1	556	139	2.3	900	95	68.34532
2	709	177	2.9	1200	135	76.27119
3	451	113	1.5	900	81	71.68142
4	587	147	2.1	900	113	76.87075
5	800	200	3.4	900	180	90
6	390	97	1	450	45	46.39175
7	401	100	1.5	450	63	63
8	306	77	0.8	450	27	35.06494
9	450	112	1.9	900	90	80.35714
10	363	91	1.4	450	31	34.06593
Total	5013	1253				

Untuk memulai tahapan proses *k-means* dilakukan *Data Selection* dari tabel 1 diatas. Kemudian dilanjutkan tahap *Pre-processing* data untuk mengubah data mentah menjadi bentuk data yang mudah dipahami dan sesuai kebutuhan. Setelah melalui tahap *pre-processing* diperoleh kondisi data sesuai dengan tabel 2 berikut ini.

Tabel 2. Hasil *Pre-processing* Data

No	Nama Dusun	Kriteria Kesejahteraan Penduduk Pondok Bungur Tahun 2023		
		Penghasilan	Listrik	% PAM
1	DUSUN 1	2300000	900	68.34
2	DUSUN 2	2900000	1200	76.27
3	DUSUN 3	1500000	900	71.68
4	DUSUN 4	2100000	900	76.87
5	DUSUN 5	3400000	900	90
6	DUSUN 6	1000000	450	46.39
7	DUSUN 7	1500000	450	63
8	DUSUN 8	850000	450	35.06
9	DUSUN 9	1900000	900	80.35
10	DUSUN 10	1400000	450	34.06

Selanjutnya masuk ke proses *data mining* dengan menggunakan metode *k-means*. Iterasi pertama dimulai dengan menentukan centroid awal yaitu centroid 1, centroid 2, dan centroid 3 dari data yang tersedia, kemudian melakukan perhitungan *k-means* pada

centroid awal tersebut. Dengan menggunakan rumus (1) perhitungan *k-means* centroid pertama pada iterasi pertama sebagai berikut:

Data Ke-1 (Dusun 1) Pada Centroid 1

$$D = \sqrt{(2300000 - 2300000)^2 + (900 - 900)^2 + (68.34 - 68.34)^2} = 0$$

Data Ke-2 (Dusun 2) Pada Centroid 1

$$D = \sqrt{(2900000 - 2300000)^2 + (1200 - 900)^2 + (76.27 - 68.34)^2} = 600000.0751$$

Data Ke-3 (Dusun 3) Pada Centroid 1

$$D = \sqrt{(1500000 - 2300000)^2 + (900 - 900)^2 + (71.68 - 68.34)^2} = 800000$$

dan seterusnya hingga data ke-10 pada dusun 10. Lakukan langkah yang sama untuk centroid 2 dan centroid 3. Sehingga diperoleh hasil perhitungan *k-means* iterasi pertama sesuai tabel 3 berikut ini.

Tabel 3. Perhitungan *K-Means* Iterasi Pertama (Minimum & Cluster)

Data Ke	C1	C2	C3	Minimum	Cluster
1	0	600000.0751	800000	0	1
2	600000.0751	0	1400000.032	0	2
3	800000	1400000.032	0	0	3
4	200000.0002	800000.0563	600000	200000.0002	1
5	1100000	500000.0902	1900000	500000.0902	2
6	1300000.078	1900000.148	500000.2031	500000.2031	3
7	800000.1266	1400000.201	450.083706	450.083706	3
8	1450000.07	2050000.138	650000.1568	650000.1568	3
9	400000.0002	1000000.045	400000.0001	400000.0001	3
10	900000.1132	1500000.188	100001.0196	100001.0196	3

Setelah hasil nilai minimum sudah ditentukan, maka tahap selanjutnya ialah menentukan data *cluster*. Dimana data *cluster* ini diperoleh dari penempatan nilai C1, C2 dan C3 yang sudah ditetapkan dinilai minimum. Pusat *cluster* baru dapat dilihat pada tabel 4 dibawah ini. Dengan pusat *cluster* baru tersebut dilakukan iterasi kedua dan seterusnya sampai tidak ada lagi perubahan pusat *cluster*. Dari pembahasan ini diperoleh iterasi keempat sebagai iterasi terakhir.

Tabel 4. Pusat *Cluster* Baru Iterasi Kedua

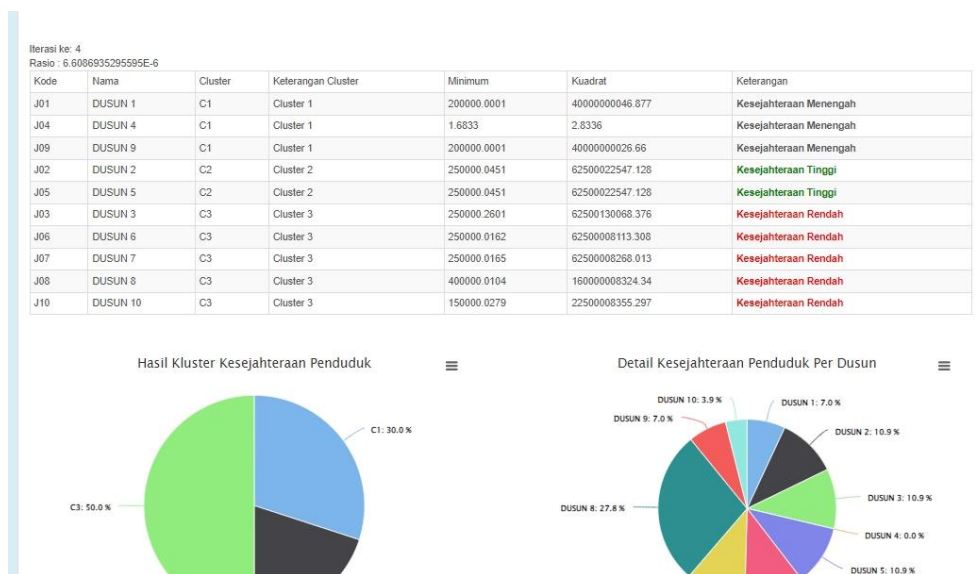
Data Ke	C1	C2	C3
1	1		
2		1	
3			1
4	1		
5		1	
6			1
7			1
8			1
9			1
10			1

Karena iterasi sudah berakhir di iterasi keempat maka dilakukan analisis akhir untuk memberikan keterangan *cluster* kepada setiap dusun, sesuai dengan hasil *cluster* terakhir yang diperoleh dari perhitungan *k-means*. Sehingga hasil dari klasifikasi kesejahteraan penduduknya seperti terlihat pada tabel 5 berikut ini.

Tabel 5. Hasil Klasifikasi *Cluster*

Data Ke	Nama Cabang	<i>Cluster</i>	Minimum	Keterangan <i>Cluster</i>
1	DUSUN 1	C1	200000.0001	Menengah
4	DUSUN 4	C1	1.683333333	Menengah
9	DUSUN 9	C1	200000.0001	Menengah
2	DUSUN 2	C2	250000.0451	Tinggi
5	DUSUN 5	C2	250000.0451	Tinggi
3	DUSUN 3	C3	250000.2601	Rendah
6	DUSUN 6	C3	250000.0162	Rendah
7	DUSUN 7	C3	250000.0165	Rendah
8	DUSUN 8	C3	400000.0104	Rendah
10	DUSUN 10	C3	150000.0279	Rendah

Hasil klasifikasi perhitungan secara manual diatas sesuai dengan hasil klasifikasi dari sistem yang dibuat (gambar 1), sehingga dapat disimpulkan sistem yang dibuat telah berjalan sesuai dengan yang diharapkan dan mampu memberikan hasil klasifikasi terhadap data penduduk untuk mengelompokan dusun di Desa Pondok Bungur sesuai kesejahteraannya sehingga mudah dalam pemberian prioritas bantuan sosial untuk kedepannya.



Gambar 1. Hasil Perhitungan Sistem

SIMPULAN

Sistem data mining yang dibuat berhasil melakukan klasifikasi dengan cepat dengan mengelompokkan penduduk menjadi 3 kluster berdasarkan dusun. Dari 10 dusun di desa Pondok Bungur, diperoleh 2 dusun dengan klasifikasi tinggi (20%), 3 dusun klasifikasi menengah (30%), dan 5 dusun yang klasifikasi masih rendah (50%) atau tingkat kesejahteraannya paling rendah dari semua dusun. Secara global tingkat kesejahteraan di Desa Pondok Bungur tergolong rendah, namun secara pemetaan desa untuk Dusun 2 dan Dusun 5 menjadi dusun dengan tingkat kesejahteraan tertinggi di desa Pondok Bungur, Dusun 1, Dusun 4, dan Dusun 9 adalah dusun-dusun yang tingkat kesejahteraan penduduknya menengah di desa Pondok Bungur. Selebihnya sisa 50% lagi masih dalam tingkat kesejahteraan rendah, sehingga dapat dijadikan sebagai prioritas dalam pemberian bantuan. Untuk perhitungan yang lebih akurat lagi disarankan dapat dikembangkan dengan metode terbaru lainnya seperti *machine learning*. Untuk keamanan sistem datamining yang dibuat disarankan dapat mengembangkannya lagi menggunakan *framework* seperti CI atau laravel. Untuk analisis yang lebih detail lagi disarankan menggunakan data penduduk yang lebih detail, atau memfokuskan penelitian untuk menyoroti pengelompokan penduduk sesuai dengan kartu keluarga nya masing-masing.

UCAPAN TERIMA KASIH

Penelitian ini dapat dilaksanakan dengan baik berkat bantuan dari berbagai pihak, untuk itu peneliti mengucapkan terima kasih kepada Kepala Desa Pondok Bungur, Kepala Dusun Pondok Bungur yang telah memberikan kerjasama yang baik dalam penelitian ini.

DAFTAR PUSTAKA

- [1] Hutabarat, L. Y., Gunawan, I., Purnamasari, I., Safii, M., & Saputra, W. (2021). Penerapan Algoritma K-Means Dalam Pengelompokan Jumlah Penduduk Berdasarkan Kelurahan Di Kota Pematangsiantar. *Jurnal Ilmu Komputer dan Teknologi*, 2(2), 20-26.
- [2] Riana, D. S. (2021). *Analisis Cluster Untuk Mengklasifikasi Tingkat Kesejahteraan Sosial Masyarakat Di Kabupaten Deli Serdang Menggunakan Metode Fuzzy C-Mean Clustering Saat Pandemi Covid-19* (Doctoral dissertation, Universitas Islam Negeri Sumatera Utara Medan).
- [3] Novitasari, N., Nuris, N. D., & Herdiana, R. (2023). Penerapan Algoritma K-Means untuk Clustering Data Jumlah Penduduk Miskin Berdasarkan Kota/Kabupaten di Jawa Barat menggunakan Rapidminer. *Jurnal Informatika Terpadu*, 9(1), 68-73.
- [4] Wakhidah, N. (2010). Clustering menggunakan k-means algorithm. *Jurnal Transformatika*, 8(1), 33-39.
- [5] Fathurrahman, F., Harini, S., & Kusumawati, R. (2023). Evaluasi clustering K-Means dan K-Medoid pada persebaran Covid-19 di Indonesia dengan metode Davies-Bouldin Index (DBI). *Jurnal Mnemonic*, 6(2), 117-128.
- [6] Indrawan, I., & Oktarina, D. (2023). Sistem Penilaian Kinerja untuk Peningkatan Akurasi Dalam Pemberian Kenaikan Gaji Karyawan dengan Metode 360 Degree. *Jurnal Mahasiswa Aplikasi Teknologi Komputer Dan Informasi (Jmapteksi)*, 4(1), 14-18.
- [7] Budiman, I., Saori, S., Anwar, R. N., Fitriani, F., & Pangestu, M. Y. (2021). Analisis Pengendalian Mutu Di Bidang Industri Makanan (Studi Kasus: Umkm Mochi Kaswari Lampion Kota Sukabumi). *Jurnal Inovasi Penelitian*, 1(10), 2185-2190.