

KLASTERISASI KARYAWAN BERDASARKAN DATA KONSULTASI PSIKOLOGI MENGGUNAKAN K-MEANS

Agid Supeno^{1*}, Rusdah²

¹Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta Indonesia

²Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta Indonesia

Email: ^{1*}1912500012@student.budiluhur.ac.id, ²rusdah@budiluhur.ac.id

(* : corresponding author)

Abstrak-Dalam psikologi adalah kemampuan individu dalam menjalani kehidupan sosial dan membangun hubungan hangat dengan orang lain, memiliki tujuan dalam menjalani hidup, mampu menerima diri, kemauan untuk terus menyalurkan potensi diri, mandiri jika dihadapkan dengan tekanan sosial, serta mampu mengontrol lingkungan eksternalnya. Faktor-faktor yang dapat memengaruhi tingkat kesejahteraan psikologi tiap individu adalah faktor demografis yang meliputi usia, status pernikahan, status sosial ekonomi dan budaya pendidikan, jenis kelamin, dan faktor dukungan sosial, religiusitas, kepribadian, serta evaluasi terhadap pengalaman. Data diperoleh dari perusahaan Xavier Psikologi kurun waktu 2021 pada penelitian ini teknik dari data mining untuk klasterisasi Kategori Masalah karyawan di Perusahaan menggunakan Algoritme *K-Means*. Dataset terdiri dari 183 data. Yaitu terdiri dari Direktorat, Jenis Kelamin, Apakah Anda Pernah Menemui Tenaga Profesional, dan Menurut Anda, Apa Kategori Masalah Tersebut dan usia. Data yang dikumpulkan akan diolah dengan menggunakan *clustering* dalam 3 *cluster* yaitu *Cluster 0* (*Cluster* rendah) terdapat 2 direktorat Finance dan *Operation* jenis kelamin Laki-laki dengan dominan Hubungan/komunikasi dengan atasan rekan kerja dengan berusia rata-rata 23 tahun. Pada *Cluster 1* (*Cluster* sedang) yang terdiri dari 1 direktorat yaitu *Technology* jenis kelamin laki-laki dengan masalah hubungan pribadi berusia rata-rata 32 tahun. Dan pada *Cluster 2* (*Cluster* Tertinggi) terdapat 2 direktorat yang terdiri dari Marketing, *operation* jenis kelamin Laki-laki dengan dominan pengembangan diri dan dengan berusia rata-rata 25 tahun.

Kata Kunci: *Data mining*, Psikologi, Algoritme *Clustering K-Means*

EMPLOYEE CLUSTERIZATION BASED ON PSYCHOLOGICAL CONSULTATION DATA USING K-MEANS

Abstract-In psychology is the individual's ability to live a social life and build warm relationships with others, have goals in living life, be able to accept oneself, the willingness to continue to channel one's potential, be independent when faced with social pressure, and be able to control the external environment. Factors that can affect the level of psychological well-being of each individual are demographic factors which include age, marital status, socio-economic status and educational culture, gender, and factors of social support, religiosity, personality, and evaluation of experience. Data were obtained from the Xavier Psychology company for the period 2021. In this study, data mining techniques for clustering employee problem categories in the company used the *K-Means* Algorithm. The dataset consists of 183 data. Which consists of the Directorate, Gender. Have you ever seen a professional, and in your opinion, what is the category of the problem, and the age. The data collected will be processed using clustering in 3 clusters, namely *Cluster 0* (low cluster) there are 2 directorates of Finance and Operations male gender with a dominant relationship / communication with superior co-workers with an average age of 23 years in *Cluster 1* (*Cluster* medium) which consists of 1 directorate namely *Technology* gender male with personal relationship problems an average age of 32 years. And in *Cluster 2* (*Highest Cluster*) there are 2 directorates consisting of Marketing, Operations Male gender with dominant self-development and with an average age of 25 years.

Keywords: *Data Mining*, Psychologist, *K-Means Clustering Algorithm*

1. PENDAHULUAN

Psikologi Industri dan Organisasi (I/O) adalah cabang psikologi yang fokus pada penerapan prinsip-prinsip psikologi dalam konteks industri dan organisasi untuk meningkatkan kinerja, efisiensi, dan kesejahteraan karyawan serta organisasi secara keseluruhan. Di bawah ini adalah penjelasan tentang bagaimana Psikologi Industri dan Organisasi dibangun[1]. Data yang digunakan mencakup beberapa variabel, seperti direktorat tempat karyawan bekerja, jenis kelamin, apakah karyawan pernah mengunjungi tenaga profesional psikologi sebelumnya, kategori masalah yang dihadapi, usia, dan hari konsultasi. Dengan menggunakan teknik *clustering k-means*,

penelitian ini berusaha untuk mengelompokkan karyawan berdasarkan pola dan karakteristik yang serupa dalam konsultasi psikologi mereka. Hal ini dapat membantu dalam mengidentifikasi masalah kesehatan mental yang mungkin dialami oleh karyawan dalam organisasi atau perusahaan tersebut. Dengan mengelompokkan data konsultasi psikologi berdasarkan variabel-variabel tersebut, penelitian ini dapat memberikan wawasan tentang kesehatan mental karyawan secara keseluruhan. Hasil dari penelitian ini dapat membantu manajemen atau departemen sumber daya manusia dalam merancang program atau intervensi yang lebih tepat dan efektif untuk mendukung kesejahteraan mental karyawan.

Untuk mencapai kesejahteraan psikologis yang baik, kondisi mental yang sehat merupakan salah satu faktor penting. Kesejahteraan psikologis melibatkan berbagai aspek dalam kehidupan seseorang yang mencakup kepuasan emosional, kognitif, dan sosial. Aspek-aspek yang telah Anda sebutkan dalam deskripsi sebelumnya, seperti kemampuan menjalani kehidupan sosial, memiliki tujuan hidup, menerima diri sendiri, menyalurkan potensi diri, mandiri di bawah tekanan sosial, dan mengontrol lingkungan eksternal, semuanya saling berhubungan dengan kesejahteraan psikologis[2]. Terdapat metode-metode dari penelitian terdahulu yang sudah melakukan *clusterisasi* pada pegawai dengan menggunakan metode K-Means dari [3],[4],[5],[6]. Selain itu, terdapat penelitian lain yang menggunakan algoritme berbedan yaitu algoritme K-Medoids dari [7].

Penelitian ini menggunakan teknik data mining dengan algoritme k-means untuk mengelompokkan data konsultasi psikologi. Algoritme k-means *clustering* dipilih karena memiliki beberapa kelebihan, yaitu mudah dipahami dan diimplementasikan, sehingga cocok untuk digunakan dalam penelitian ini. Selain itu, algoritme ini juga memiliki waktu pemrosesan yang relatif cepat, sehingga memungkinkan untuk mempelajari data dengan ukuran yang cukup besar.

2. METODE PENELITIAN

2.1 Metode CRISP-DM

Untuk data mining yang sebenarnya populer dan sering digunakan adalah "CRISP-DM" (Cross-Industry Standard Process for Data Mining), **bukan** "Cross Standard Industry for Data Mining". CRISP-DM adalah model proses standar yang terdiri dari enam fase untuk membimbing proses data mining dari awal hingga akhir [8]. Gambar 1 menggambarkan proses CRISP-DM.



Gambar 1. Proses CRISP-DM [8]

2.1.1 *Understanding Business/Project Understanding*

Fase pertama ini berfokus pada memahami tujuan bisnis atau proyek dan menetapkan tujuan yang spesifik untuk proses data mining. Identifikasi masalah yang ingin dipecahkan dan kebutuhan informasi yang ingin dipenuhi.

2.1.2 *Data Understanding*

Pada tahap ini, Anda akan mengumpulkan data yang relevan untuk analisis, mengeksplorasi data tersebut, dan melakukan analisis awal untuk memahami karakteristiknya. Anda juga akan mengidentifikasi potensi masalah data seperti kehilangan data, outlier, atau kesalahan.

2.1.3 *Data Preparation*

Tahap ini melibatkan pembersihan data, integrasi data dari berbagai sumber, pemilihan atribut (fitur) yang relevan, dan transformasi data untuk mempersiapkan data mentah menjadi data yang siap untuk proses data mining.

2.1.4 *Modeling*

Pada tahap ini, Anda akan memilih dan menerapkan algoritme data mining yang sesuai untuk menciptakan model prediktif atau deskriptif. Model ini akan didasarkan pada data yang telah dipersiapkan pada tahap sebelumnya.

2.1.5 *Evaluation*

Evaluasi model dilakukan untuk mengukur kinerja dan efektivitasnya dalam mencapai tujuan bisnis atau proyek. Berbagai metrik evaluasi digunakan untuk menilai seberapa baik model bekerja.

2.1.6 *Deployment*

Tahap terakhir adalah implementasi model dalam produksi atau penggunaan sehari-hari. Model yang telah diuji dan dievaluasi dengan baik dapat diterapkan untuk mengambil keputusan atau menyediakan solusi untuk masalah yang ditentukan sebelumnya.

Setelah tahapan *deployment*, proses data mining biasanya juga melibatkan tahap monitoring dan maintenance untuk memastikan bahwa model tetap berfungsi dengan baik seiring waktu dan mengikuti perubahan dalam lingkungan bisnis. CRISP-DM menyediakan kerangka kerja yang sistematis dan terstruktur untuk proyek data mining, memungkinkan analisis data untuk mengatasi tantangan dan permasalahan yang sering muncul dalam proses analisis data yang kompleks.

2.2 *Metode Elbow*

Metode Elbow adalah salah satu teknik yang digunakan dalam analisis klusterisasi untuk membantu menentukan jumlah kluster yang optimal dalam kumpulan data. Tujuannya adalah untuk menemukan titik di mana penambahan kluster tidak lagi memberikan penurunan yang signifikan dalam variabilitas data [9].

2.3 *Algoritme K-Means*

K-Means *Clustering* merupakan salah satu algoritme *Clustering* yang Masuk ke dalam kelompok Unsupervised learning yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok dengan sistem partisi. Algoritme data mining untuk mencari pola dari semua variabel, variabel yang menjadi target/label/class tidak ditentukan. Algoritme K-Means merupakan teknik *Clustering* berbasis jarak yang Membagi data ke dalam beberapa *cluster* dan algoritme ini hanya Bekerja pada atribut angka dan numerik [10].

3. HASIL DAN PEMBAHASAN

3.1. *Business Understanding*

Penelitian melakukan pemahaman terhadap kebutuhan data terkait pegawai mengikuti konsultasi psikologi yang telah diadakan oleh perusahaan. Penelitian ini menggunakan data dari pegawai yang telah mengikuti tes psikologi yang di laksanakan pada tahun 2022. Data ini meliputi 183 pegawai yang diliputi data: Direktorat, jenis kelamin, apakah Anda pernah menemui tenaga profesional, menurut Anda apa kategori masalah tersebut. Data tersebut didapatkan dari total seluruh data pegawai di perusahaan.

3.2. Data Preparation

Mempersiapkan data dilakukan beberapa proses diantaranya pemilihan variabel untuk dianalisis, pembersihan data, persiapan data. Data selection yang di ambil di suatu kantor psikologi. Selanjutnya, melakukan data preprocessing. Pada proses ini bertujuan untuk melihat kualitas data yang dipilih pada proses ini bertujuan untuk melihat kualitas data yang dipilih pada proses data selection. Pada tahap ini dilakukan pembersihan data (Data *cleansing*) yaitu proses pembersihan data dari *missing value* atau *noise*. Tahap berikutnya adalah tranformasi data dengan cara normalisasi agar tidak ada perbedaan yang terlalu jauh pada skala data.

3.2.1. Data Transformation

Pada tahapan ini untuk mengubah seluruh data yang terkumpul agar memudahkan ketika diolah dengan menggunakan metode K-Means *clustering*. Adapun variabel yang digunakan pada penelitian ini yaitu Direktorat, Jenis kelamin, Apakah anda pernah menemui tenaga profesional, Kategori masalah tersebut, dan Usia. Berikut beberapa atribut yang di transformasi dengan label *encoding*:

- a. Direktorat: Hasil Transformasi Data Direktorat dapat dilihat pada Tabel 1.

Tabel 1. Hasil Transformasi Data Direktorat

Direktorat	Kode
<i>Operation</i>	111
<i>CEO Office</i>	222
<i>Marketing</i>	333
<i>Finance</i>	444
<i>Technology</i>	555

- b. Jenis kelamin: Hasil Transformasi Data Jenis kelamin dapat dilihat pada Tabel 2.

Tabel 1. Hasil Transformasi Data Jenis kelamin

Jenis Kelamin	Kode
Laki-Laki	10
Perempuan	20

- c. Apakah anda pernah menemui tenaga profesional: Hasil Transformasi Data Apakah anda pernah menemui tenaga profesioanal dapat dilihat pada Tabel 3.

Tabel 2. Hasil Transformasi Data Apakah Anda Pernah Menemui Tenaga Profesional

Apakah Anda Pernah Menemui Tenaga Profesional	Kode
Belum Pernah	100
Psikologi	200
Psikiater	300

- d. Kategori masalah tersebut: Hasil Transformasi Data Kategori Masalah dapat dilihat pada Tabel 4.

Tabel 3. Hasil Transformasi Data Kategori Masalah

Kategori Masalah tersebut	Kode
Pengembangan diri	500
Masalah keluarga	600
Masalah Hubungan Pribadi	700
Keseimbangan Antara Pekerjaan dan Kehidupan Pribadi	800
Hubungan/Komunikasi Dengan Atasan dan Rekan Kerja	900
Stres/Kelelahan Terhadap Pekerjaan	910
Kecemasan (Perasaan takut, gugup, atau gelisah yang berlebihan)	920
Kurang Motivasi	930
Beradaptasi Dengan Perubahan di Kehidupan	940

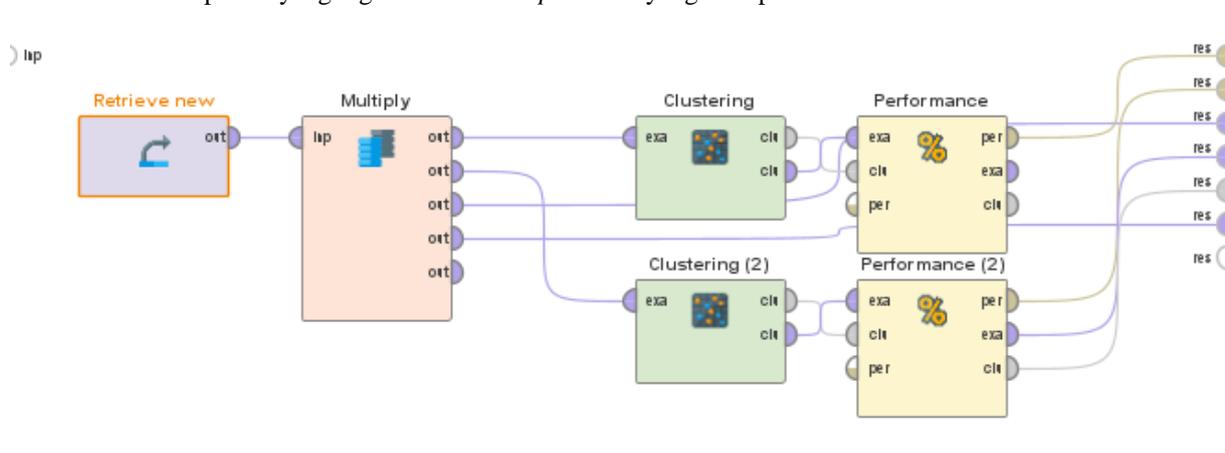
e. Usia: Hasil Transformasi Data Usia dapat dilihat pada Tabel 5.

Tabel 4. Hasil Transformasi Data Usia

Hasil Pembentukan Atribut Usia
34
27
29
28
24
32
26
38
31
33
41
23
25
30
29
28
24
43
37

3.3. Modelling

Pada tahap ini penelitian melakukan tahap modeling untuk menentukan *cluster* yang optimal menggunakan k-means. Berikut proses yang digunakan oleh *RapidMiner* yang terdapat di Gambar 2.



Gambar 2. Proses dari Algoritme K-Means

Setelah melakukan proses *RapidMiner*, maka dapat dilihat hasil pada tabel 6 nilai DBI paling kecil dengan menentukan jumlah *cluster* terbaik adalah berjumlah 3 *cluster* dengan nilai 0,717. Semakin kecil nilai DBI maka semakin optimal *cluster* yang dihasilkan.

Tabel 5. Hasil Davies Bouldin Indeks

Penentuan Jumlah <i>Cluster</i> Terbaik	Hasil Davis Bouldin Indeks
2 <i>Cluster</i>	0.898
3 <i>Cluster</i>	0.717
4 <i>Cluster</i>	0.726

Selanjutnya, gambar 3 merupakan hasil 3 *cluster* yang diperoleh.

Cluster Model

```
Cluster 0: 76 items
Cluster 1: 47 items
Cluster 2: 59 items
Total number of items: 182
```

Gambar 3. Hasil Jumlah *Cluster*

Dengan menggunakan pemodelan K-Means *clustering* seperti pada gambar 3, dengan jumlah data 182 dan inisialisasi jumlah *cluster* sebanyak 3 buah, sesuai dengan pendefinisian nilai k dengan jumlah *cluster_0* : 76 items, *cluster_1* : 47 items, *cluster_2* : 59 items.

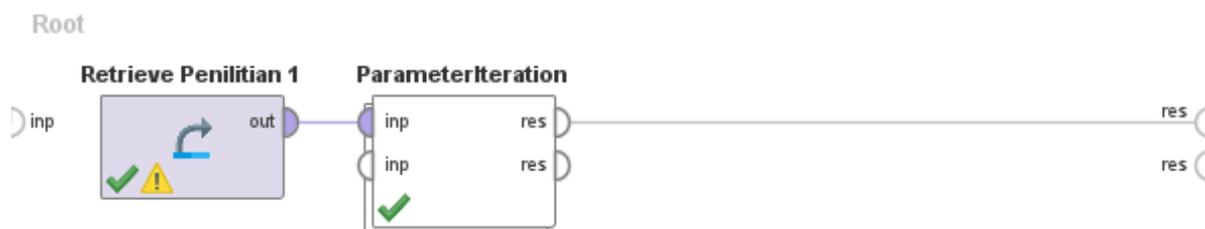
Dari data hasil penelitian *clustering* yang telah dilakukan, dapat disimpulkan pada *cluster 0* merupakan *Cluster* rendah dengan Keluhan yang di alami adalah Masalah Hubungan Pribadi, Beradaptasi dengan perubahan di kehidupan, Hubungan atau komunikasi dengan atasan dan rekan kerja, Kecemasan (Perasaan takut, gugup, atau gelisah yang berlebihan), Keseimbangan antara pekerjaan dan kehidupan pribadi, Kurang Motivasi Kerja, Stres / kelelahan terhadap pekerjaan tetapi untuk keluhan yang dominan ialah Hubungan/Komunikasi dengan atasan dan rekan kerja, dengan dominan Direktorat finance dan opration, dan dominan Jenis kelamin laki-laki, serta belum pernah menemui tenaga profesional dominan pernah menemui yaitu Psikologi serta berumur 23 dan 27.

Pada *cluster 1* merupakan *cluster* sedang dengan keluhan yang di alami adalah Masalah Hubungan Pribadi, Masalah Keluarga, Pengembangan diri, untuk keluhan yang dominan ialah Masalah Hubungan Pribadi, dengan dominan Direktorat Technolgy dan dominan Jenis Kelamin Laki-laki serta belum pernah menemui tenaga Profesional dominan Belum Pernah serta berumur 32.

Selanjutnya *Cluster 2* merupakan *cluster* tinggi dengan keluhan yang di alami adalah Keseimbangan antara pekerjaan dan kehidupan pribadi, Masalah Hubungan Pribadi, Pengembangan diri, Masalah Keluarga, untuk keluhan yang dominan ialah Pengembangan diri, dengan dominan Direktorat Marketing dan dominan Jenis kelamin Laki-laki, serta belum pernah menemui tenaga Profesional dominan Belum Pernah serta berumur 25.

3.4. Evaluation

Pada bagian ini disampaikan mekanisme pengukuran perfoeme model terbaik. Metode uji dari penelitian ini menggunakan *Davies Bouldin Indeks*. Berikut proses dari pengujian yang menggunakan *tools RapidMiner* dapat dilihat pada Gambar 4:



Gambar 4. Operator Pengujian

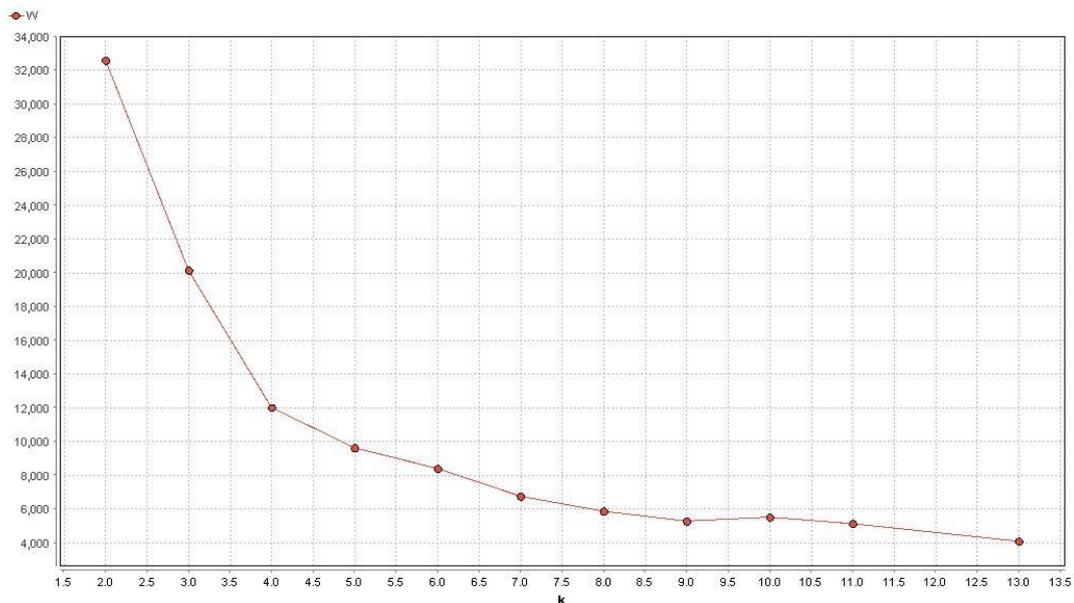
Proses Gambar 4 di atas menggunakan operator loop parameters. Lalu dari hasil processLog menunjukkan nilai DB terkecil ialah 0,717 dengan jumlah 3 *cluster* yang optimal. Gambar 5 menunjukkan hasil pengujian dengan nilai DB terkecil.

ProcessLog (11 rows, 3 columns)

k	DB	W
2	0.898	32379.025
3	0.717	20077.100

Gambar 5. Hasil Pengujian

Dalam kasus ini, karena terdapat penurunan yang signifikan pada nilai DBI untuk *cluster* 3.0, dan di antara titik tersebut terdapat "siku" yang mencerminkan perubahan penurunan yang lebih lambat, maka *cluster* dengan jumlah 3 dapat dianggap sebagai kandidat *cluster* optimal. Hasil grafik dapat dilihat pada Gambar 6.



Gambar 6. Hasil Grafik

4. KESIMPULAN

Berdasarkan hasil penelitian dengan bantuan tools RapidMiner menggunakan algoritme *k-means* didapatkan nilai Davies Boulding Index paling kecil dengan penentuan jumlah *cluster* terbaik adalah berjumlah 3 *cluster* yaitu 0.717. Berdasarkan hasil penelitian yang ada dapat disimpulkan bahwa terdapat 3 *cluster*, yaitu *Cluster* 0 (Tingkat karyawan rendah) terdapat 2 direktorat Finance dan Opration jenis kelamin Laki-laki dengan dominan Hubungan/komunikasi dengan atasan rekan kerja dengan berusia rata-rata 23 tahun. *cluster* 1 (Tingkat karyawan sedang) yang terdiri dari 1 direktorat yaitu Technolgy jenis kelamin laki-laki dengan masalah hubungan pribadi berusia rata-rata 32 tahun. *cluster* 2 (Tingkat karyawan tinggi) terdapat 2 direktorat yang terdiri dari *Marketing, operation* jenis kelamin Laki-laki dengan dominan pengembangan diri dan dengan berusia rata-rata 25 tahun.

DAFTAR PUSTAKA

- [1] Umama, H. A. (2019). *Buku ajar psikologi industri dan organisasi*.
- [2] Izzati, U. A., Budiani, M. S., Mulyana, O. P., & Puspitadewi, N. W. S. (2021). Gambaran Kesejahteraan Psikologis pada Karyawan Terdampak Pandemi COVID-19. *Jurnal Psikologi Teori dan Terapan*, 11(3), 315-325.
- [3] Setiawan, S. (2019). Analisis Cluster Menggunakan Algoritme K-Means Untuk Mengetahui Kemampuan Pegawai Dibidang It Pada Cv. Roxed Ltd. *Pelita Informatika: Informasi dan Informatika*, 7(3), 341-347.
- [4] Kristanto, B., Zy, A. T., & Fatchan, M. (2023). Analisis Penentuan Karyawan Tetap Dengan Algoritme K-Means Dan Davies Bouldin Index. *Bulletin of Information Technology (BIT)*, 4(1), 112-120. (2002) The IEEE website. [Online]. Available: <http://www.ieee.org/>
- [5] Zuraidah, E., & Yuliyanti, I. D. (2021). Penerapan Sistem Pendukung Keputusan Penerimaan Karyawan Baru Di PT. Patra Jasa Dengan Metode K-Means Clustering. *Resolusi: Rekayasa Teknik Informatika dan Informasi*, 2(1), 7-14.
- [6] Intan, M., Priyatna, B., Nurapriani, F., & Hilabi, S. S. (2023). Implementasi Metode K-Means Clustering Pada Penilaian Kinerja Karyawan PT Kopetri Citra Abadi: Metode K-Means Clustering. *Informasi Interaktif: Jurnal Informatika dan Teknologi Informasi*, 8(1), 24-30.
- [7] Kaligis, G. B., & Yulianto, S. (2022). Analisa Perbandingan Algoritme K-Means, K-Medoids, Dan X-Means Untuk Pengelompokkan Kinerja Pegawai. *IT-EXPLORE: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, 1(3), 179-193.
- [8] Suhanda, Y., Kurniati, I., & Norma, S. (2020). Penerapan Metode Crisp-DM Dengan Algoritme K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika dan Komputer*, 6(2), 12-20.
- [9] Dewi, D. A. I. C., & Pramita, D. A. K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritme Clustering KMedoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix: Jurnal Manajemen Teknologi dan Informatika*, 9(3), 102-109.
- [10] Nabila, Z., Isnain, A. R., Permata, P., & Abidin, Z. (2021). Analisis data mining untuk clustering kasus covid-19 di Provinsi Lampung dengan algoritme k-means. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 100-108.