PENERAPAN METODE K-MEANS CLUSTERING PADA DATA PENJUALAN OPTIK CHANTIKA

Muhamad Rizki, Mulyawan
Program Studi Komputerisasi Akuntansi D3
STMIK IKMI Cirebon, Jl. Perjuangan No. 10 B Majasem Kota Cirebon, Indonesia
mr1188027@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk melakukan analisis data mining menggunakan metode clustering K-Means pada data penjualan merek kacamata di Optik Chantika. Dalam penelitian ini, kami menggunakan nilai Davies Bouldin Index untuk mengevaluasi performa metode K-Means, dan hasil optimal terdapat pada jumlah klaster 10. Namun, kami memilih menggunakan tiga klaster untuk klasterisasi pada penelitian ini: penjualan tertinggi, penjualan sedang dan penjualan terendah. Melalui metode K-Means, merek kacamata dengan penjualan tertinggi selama tiga bulan tergabung dalam Cluster_0, sedangkan Cluster_1 berisi merek kacamata dengan penjualan sedang, dan Cluster_2 berisi merek kacamata dengan penjualan terendah. Hasil penelitian ini memberikan persentase 50% merek kacamata termasuk dalam kategori penjualan tertinggi, 27,27% termasuk dalam kategori penjualan sedang, dan 22,72% termasuk dalam kategori penjualan terendah. Dengan adanya pengelompokan berdasarkan pola penjualan ini, Optik Chantika dapat mengoptimalkan pengelolaan data penjualan dan memenuhi kebutuhan konsumen dengan lebih baik. Dengan pemahaman yang lebih baik tentang pola penjualan, perusahaan dapat mengambil keputusan yang lebih efektif dalam hal pengadaan stok dan strategi pemasaran. Hal ini diharapkan dapat meningkatkan pendapatan perusahaan dan memberikan manfaat yang signifikan bagi Optik Chantika.

Kata kunci : K-Means, Optik Chantika, Clustering

1. PENDAHULUAN

Persaingan dalam memasarkan produk atau jasa menjadi semakin ketat terutama dalam era globalisasi yang kompetitif. Hal ini mendorong para pengusaha untuk melakukan kegiatan pemasaran yang lebih efektif dan berfokus pada kebutuhan konsumen. Perubahan selera konsumen dan perubahan di lingkungan sekitar telah mengubah lanskap bisnis yang semakin dinamis. Pertumbuhan kebutuhan konsumen yang meningkat juga membuka peluang bisnis yang menguntungkan. Salah satu jenis bisnis yang banyak dibutuhkan oleh masyarakat dari berbagai kalangan dan usia adalah toko kacamata, yang umumnya dikenal sebagai optik. Optik Chantika merupakan salah satu toko penjualan alat-alat yang membantu memperbaiki gangguan penglihatan pada mata seperti lensa kontak dan kacamata dengan berbagai merek dan ukuran sesuai kebutuhan konsumen. Namun Optik Chantika belum bisa menentukan merek bingkai kacamata mana yang terlaris dan mana yang kurang diminati sehingga berakibat pada kurang tepatnya ketersediaan kacamata yang sesuai dengan kebutuhan konsumen. Maka dari itu peneliti berencana untuk melakukan data mining dengan teknik klasterisasi terhadap data-data penjualan Optik Chantika agar dapat mengatasi masalah yang ada.

Penjualan dapat diartikan sebagai kegiatan menjual barang dagang yang dilakukan secara terus menerus, sehingga penjualan menjadi salah satu langkah pemasaran yang harus dilakukan pengusaha untuk dapat memperoleh keuntungan agar kegiatan operasional perusahaan dapat terus dilakukan. *Data mining* adalah suatu proses yang dilakukan dengan tujuan untuk menemukan hubungan dari suatu data yang belum diketahui oleh pengguna, dengan menyajikannya menggunakan cara yang lebih mudah untuk dipahami agar dapat menjadi dasar pengambilan keputusan dalam penjualan barang dagang [1].

Penerapan data mining pada data penjualan merek kacamata Optik Chantika dilakukan menggunakan teknik klasterisasi dengan menggunakan metode algoritma K-means pada perangkat lunak Rapidminer. Klasterisasi dalam data mining adalah suatu teknik yang digunakan untuk mengelompokkan data ke dalam kelompok tertentu, sehingga data dalam setiap kelompok memiliki kesamaan dan perbedaan yang jelas dengan data pada kelompok lainnya [2]. Algoritma K-means merupakan salah satu teknik dalam data mining yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan jarak, kriteria, kondisi, atau karakteristik. Tujuan akhir dari algoritma K-means adalah agar data dalam setiap kelompok memiliki jarak terpendek, kriteria, kondisi, atau karakteristik yang serupa atau hampir serupa satu sama lain. Oleh karena itu, dalam penelitian ini, algoritma K-means digunakan untuk mengelompokkan objek-objek yang memiliki kemiripan [3].

Banyaknya jenis merek kacamata berbeda yang terjual dalam setiap bulan dan data penjualan yang hanya diarsipkan dan belum dikelola dengan baik

menyebabkan seringnya terjadi kekurangan stok merek kacamata yang paling tinggi penjualannya, sehingga penelitian tugas akhir ini bertujuan untuk melakukan data mining dengan teknik klasterisasi terhadap data-data penjualan Optik Chantika agar dapat mengatasi masalah yang ada. Dengan melakukan klasterisasi pemilik usaha Optik Chantika dapat memprioritaskan merek kacamata yang paling banyak dijual agar peristiwa kekurangan stok merek kacamata tidak terjadi lagi dan dapat memutuskan strategi penjualan terbaik berdasarkan data yang ada di masa depan.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining menjadi sangat signifikan karena jumlah data yang disimpan dalam basis data di berbagai bidang seperti bisnis semakin meningkat. Proses data mining, juga dikenal sebagai KDD (knowledge discovery in database), bertujuan untuk mengeksplorasi data dalam basis data dan menghasilkan informasi baru yang berharga. Selain itu, data mining juga dikenal sebagai pengenalan pola karena tujuannya adalah untuk mengidentifikasi pola yang ada dalam basis data. Oleh karena itu, data mining dapat dijelaskan sebagai proses pengolahan dan identifikasi informasi yang berguna melalui penggunaan teknik dari bidang statistik, matematika, kecerdasan buatan, dan pembelajaran mesin pada himpunan data yang besar [1].

2.2. Metode Klasterisasi

Salah satu teknik yang diterapkan dalam data mining adalah klasterisasi, yang mengacu pada metode untuk mengelompokkan data ke dalam klaster tertentu. Tujuan klasterisasi adalah agar data dalam setiap klaster memiliki kesamaan dan perbedaan yang jelas dibandingkan dengan data pada klaster lainnya. Setiap klaster yang terbentuk merupakan hasil dari identifikasi pola dalam pengelompokan data, sehingga hasil klasterisasi tersebut dapat digunakan sebagai pendukung dalam pengambilan keputusan bagi pihak yang membutuhkannya [2].

Metode klasterisasi memiliki tujuan untuk mengidentifikasi kelompok objek yang serupa di dalam satu kelompok, namun berbeda dengan objek di kelompok lainnya. Metode ini digunakan untuk mengelompokkan kasus berdasarkan kelompok atribut, dengan maksud untuk mengenali kelompok-kelompok alami. Klasterisasi adalah salah satu metode dalam data mining yang bersifat unsupervised, karena tidak ada atribut yang digunakan sebagai panduan dalam prosesnya, sehingga semua atribut input diperlakukan dengan cara yang sama. Algoritma ini membangun sebuah model dengan melakukan iterasi berulang, dan berhenti ketika model tersebut telah mencapai titik pusat atau konvergensi, yang menandakan bahwa batasan segmentasi telah stabil [4].

2.3. Algoritma K-Means

Algoritma K-Means adalah sebuah metode yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan jarak, kriteria, kondisi, atau karakteristik tertentu. Data dalam satu kelompok diharapkan memiliki jarak terpendek dan memiliki kesamaan atau kemiripan dalam kriteria, kondisi, atau karakteristik. Dengan demikian, algoritma K-Means memungkinkan pengelompokkan objek yang memiliki kesamaan di antara mereka [3].

K-Means merupakan algoritma heuristik yang digunakan untuk membagi data ke dalam kelompok K dengan cara meminimalkan total jarak kuadrat di setiap kelompok. Metode ini juga merupakan salah satu pendekatan non-hierarkis dalam pengelompokan data yang bertujuan untuk mempartisi data ke dalam satu atau lebih kluster atau kelompok. Dalam metode ini, data yang memiliki karakteristik serupa atau mirip akan dikelompokkan bersama dalam satu kluster, sedangkan data dengan karakteristik yang berbeda akan dikelompokkan ke dalam kelompok yang berbeda. Berikut adalah langkah-langkah yang ditempuh dalam algoritma K-Means:

1. Menentukan jumlah kluster yang diinginkan,
2. Menempatkan data yang ada kedalam suatu kluster secara acak dengan rumus berikut:
   $V_{ij} = \frac{1}{Ni} \sum_{K-0}^{Ni} Xkj$
   Keterangan:
   V_{ij} = nilai rata – rata centroid pada Cluster ke – i untuk variabel ke –j
   Ni = Jumlah anggota Cluster ke –i
   i,k = Indeks dari Cluster
   j = Indeks variabel
   Xkj = Nilai data ke –k variabel ke – j untuk Cluster tersebut,
3. Menghitung centroid atau rata-rata dari data yang sudah ada dari setiap kluster dengan rumus berikut:
   $D(x_2, x_1) = \| X_2 - X_1 \| 2$
   Keterangan :
   D = Euclidean Distance
   X = Banyaknya objek
   $De = \sqrt{(xi - si)^2 + (yi - ti)^2}$
   Keterangan :
   De = Euclidean Distance
   I = Banyaknya Objek2
   (x,y) = Koordinat Objek
   (s,t) = Koordinat Centroid
4. Melakukan pengelompokkan berdasarkan centroid terdekat.
5. Mengulangi langkah ke-2, dan melanjutkan iterasi hingga centroid mencapai nilai yang optimal [3].

2.4. *Rapidminer Tool*

RapidMiner merupakan sebuah perangkat lunak sumber terbuka yang digunakan untuk memproses analisis data dan data mining. Alasan utama penggunaan *RapidMiner* adalah karena menyediakan

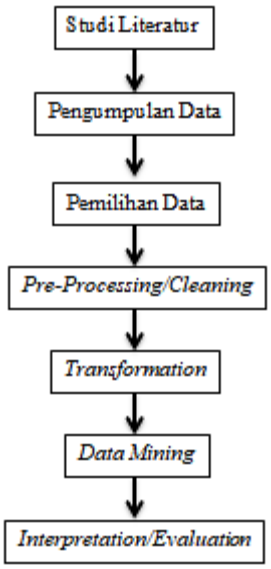
aplikasi mandiri yang mampu melakukan berbagai tugas dalam analisis data dan berfungsi sebagai mesin data mining. Perangkat ini memungkinkan pengguna untuk melakukan proses seperti memuat data, transformasi data, pemodelan data, dan visualisasi data dengan berbagai metode yang tersedia [11].

2.5. Profil Perusahaan

Optik Chantika adalah toko kacamata yang berlokasi di Jalan Bahagia No.51, Panjunan, Lemahwungkuk,Kota Cirebon dan telah berdiri sejak tahun 2008. Optik Chantika menyediakan berbagai macam merek frame dan ukuran kacamata sesuai kebutuhan konsumennya. Berdasarkan hasil observasi yang telah penulis lakukan pada Optik Chantika, diperoleh dataset penjualan merek kacamata bulan Januari sampai bulan Maret tahun 2022. Dataset tersebut yang akan diolah dalam penelitian ini.

3. METODE PENELITIAN

Alur penelitian ini digambarkan pada gambar dibawah:



Gambar 1. Alur penelitian

3.1. Studi Literatur

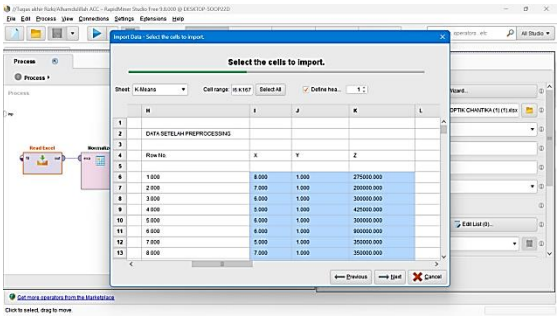
Studi literatur dilakukan untuk mencari informasi yang relevan dengan teori, metode, dan konsep yang sesuai dengan permasalahan yang ada dalam penelitian ini. Sumber informasi yang digunakan dapat berupa jurnal, e-book, literatur, internet, atau sumber lainnya yang berkaitan.

3.2. Pengumpulan Data

Data dikumpulkan dengan melakukan wawancara pada pegawai toko dan dokumentasi dokumen-dokumen penjualan kacamata Optik Chantika.

3.3. Pemilihan Data

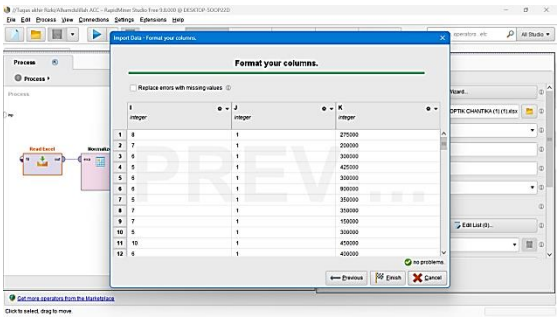
Sebelum memulai tahap KDD (Knowledge Discovery in Databases), diperlukan proses seleksi data. Data yang telah melalui proses seleksi ini akan digunakan dalam proses data mining dan akan disimpan dalam sebuah file terpisah dari basis data operasional [5]. Data yang dipilih pada penelitian ini adalah data penjualan kacamata dari Optik Chantika dari bulan Januari sampai Maret 2022. Data penjualan kacamata yang diterima dari Optik Chantika selanjutnya diperiksa dan disesuaikan formatnya dengan yang dibutuhkan. Data ini disimpan dalam format .xlsx (excel) terlebih dahulu dan kemudian diimport ke dalam *Rapidminer tool* seperti pada gambar 2. Data terpilih selanjutnya memasuki tahap *Pre-processing* atau *cleaning.*



Gambar 2. Pemilihan data penjualan kacamata optik chantika

3.4. *Pre-Processing/Cleaning*

Pemeriksaan *missing value* terhadap data yang tidak konsisten berdasarkan rata-rata data dan perbaikan data dengan format yang telah ditentukan dilakukan pada tahap ini, sehingga data yang akan digunakan nantinya menjadi bersih [6]. Data terpilih selanjutnya dibersihkan dengan memastikan tidak ada data yang hilang, tidak valid ataupun salah ketik pada tahap ini. Gambar 3. menunjukkan bahwa pada data penjualan kacamata Optik Chantika ini 4 features dengan tipe numerik (yaitu ID, QTY Januari, QTY Februari dan QTY Maret) serta data yang digunakan telah dipastikan *no missing value*. Data yang telah lolos tahap ini selanjutnya memasuki tahap *Transformation.*



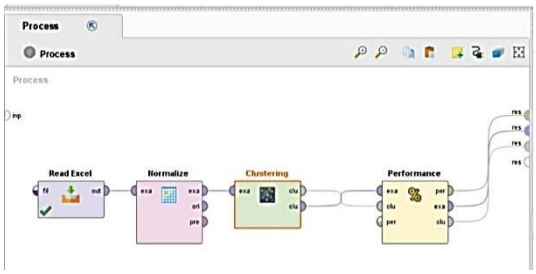
Gambar 3. *Pre-processing* data penjualan kacamata optik chantika

3.5. Transformation

Transformasi data dilakukan untuk mengubah data dengan nilai normal menjadi bentuk numerik [7]. Dalam proses ini, data yang telah terpilih akan diubah menjadi prosedur data mining, yaitu proses yang bergantung pada jenis atau pola yang ingin ditemukan dalam basis data [8]. Data yang ada sudah berbentuk numerik atau angka, sehingga tidak ada yang harus diubah dan dapat langsung digunakan ke proses berikutnya yaitu data mining.

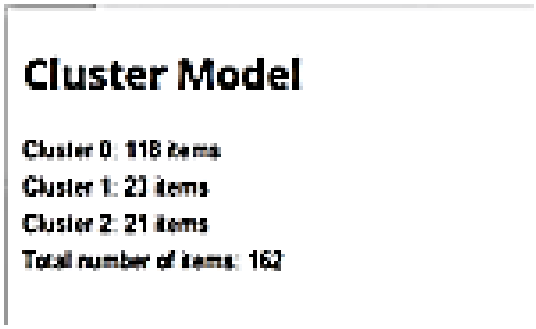
3.6. Data Mining

Proses *data mining* dilakukan untuk menetapkan algoritma atau proses pencarian pola atau informasi yang menarik dalam data terpilih. Pemilihan algoritma tergantung pada tujuan serta pada proses KDD secara keseluruhan [9]. Pada penelitian ini untuk memudahkan proses data mining maka peneliti menggunakan aplikasi Rapidminer dengan algoritma K-means. Gambar 4. menunjukkan desain widget yang digunakan pada pengolahan data mining dengan *Rapidminer tool* di penelitian ini.



Gambar 4. Widget Rapidminer Data Mining

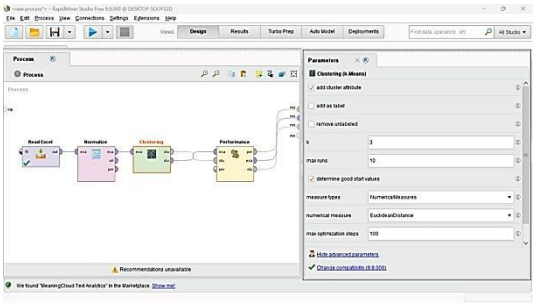
Sedangkan gambar 5. dibawah menunjukkan hasil klasterisasi dengan algoritma K-Means yang berarti bahwa 162 merek kacamata tersebut akan dikelompokkan menjadi 3 klaster yaitu cluster 0, cluster 1 dan cluster 2. Cluster 0 berjumlah 118 items, cluster 1 berjumlah 23 items, dan cluster 2 berjumlah 21 items.



Gambar 5. Desain klasterisasi dengan algoritma k-means

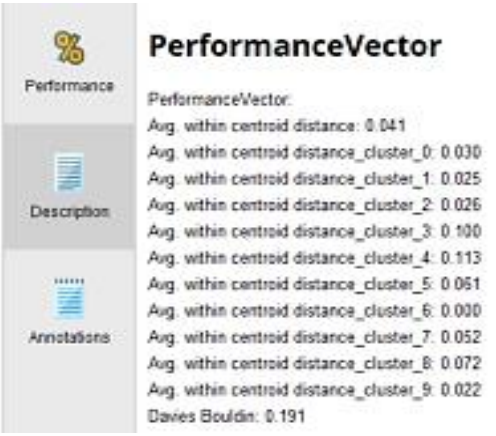
3.7. Interpretation/Evaluation

Tujuan dari tahap ini adalah melakukan evaluasi terhadap hasil dari proses data mining dan memastikan apakah hasil tersebut dapat memenuhi tujuan yang telah ditetapkan. Evaluasi ini melibatkan profilisasi dari setiap klaster yang terbentuk dan analisis lebih lanjut untuk menghubungkannya dengan atribut yang menarik. Pola-pola informasi yang ditemukan akan disajikan dalam bentuk yang mudah dipahami oleh pihak-pihak yang berkepentingan dalam proses data mining [10]. Pada penelitian ini tahap evaluasi dilakukan dengan melihat nilai Davies Bouldin Index menggunakan model performance. Nilai k yang optimum adalah nilai k yang Davies Bouldin Indexnya mendekati 0. Fungsi Operator Performance ini adalah untuk mengetahui hasil DBI menggunakan nilai k yang ditentukan secara acak, dan dilakukan beberapa kali percobaan dengan tujuan mengetahui nilai k yang paling mendekati 0. Gambar 6. dibawah menunjukkan pengaturan parameter k-mean yang digunakan untuk mencari jumlah klaster optimal pada penelitian ini.



Gambar 6. Pengaturan Parameter K-Mean

Untuk mengetahui hasil Davies Bouldin Index yang optimal dilakukan beberapa kali percobaan dengan nilai k yang berbeda. Peneliti melakukan 9x percobaan dengan nilai k yang berbeda dan diketahui nilai DBI yang paling optimal yaitu 0,191 dengan nilai k = 10 seperti yang ditunjukkan pada gambar 7. dibawah.



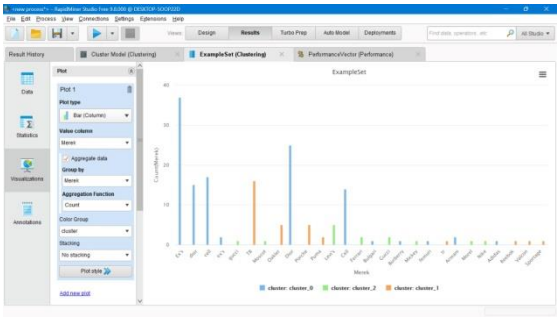
Gambar 7. Nilai Davies Bouldin Index Paling Optimal

Nilai Davies Bouldin Index hasil beberapa kali percobaan untuk mencari jumlah klaster paling optimal dapat dilihat pada tabel 4.1 dibawah ini.

Tabel 1. Percobaan Mencari Nilai Davies Bouldin Index yang Optimal

Jumlah k	Nilai DBI
2	0.292
3	0.234
4	0.230
5	0.253
6	0.223
7	0.234
8	0.218
9	0.196
10	0.191

4. HASIL DAN PEMBAHASAN



Gambar 8. Hasil klasterisasi data penjualan kacamata optik chantika

Hasil data mining dengan metode K-Means menggunakan aplikasi Rapidminer menunjukkan hasil interpretasi atau evaluasi dari data yang telah diolah dalam bentuk *scatter* plot. Axis x merupakan merek kacamata yang terjual berjumlah 22. Sementara Axis y merupakan *count (merek)* atau banyaknya jumlah kacamata yang terjual dari masing-masing merek. Bar berwarna biru menunjukkan cluster_0, bar berwarna hijau menunjukkan cluster_1 dan bar berwarna orange menunjukkan cluster_2. Gambar 8. menunjukkan bahwa dari 22 merek kacamata yang dikelompokkan ke dalam 3 klaster, yaitu cluster_0, cluster_1 dan cluster_2. Cluster_0 berisi 11 merek kacamata dengan penjualan terbanyak yaitu Ex's, Dior, Cell, Bulgari, Femuri, Ferrari, Gucci, Levi's, Mickey, Armani dan Adidas. Cluster_1 berisi 6 merek dengan penjualan sedang yaitu Oakler, Porche, Puma, Reebok, Burberry dan Sportage. Dan cluster_2 berisi 5 merek kacamata dengan penjualan rendah yaitu TR, Moscot, Valcon, Morel dan Nike.

5. KESIMPULAN DAN SARAN

Klasterisasi penjualan merek kacamata menggunakan metode K-means menghasilkan 3 kluster, yaitu cluster_0, cluster_1 dan cluster_2. Merek kacamata dengan penjualan paling banyak selama 3 bulan dikelompokkan dalam cluster_0 yang berjumlah 11 merek. Cluster_1 berisi 6 merek kacamata dengan penjualan sedang dan cluster_2 berisi 5 merek kacamata dengan penjualan rendah. Tingkat persentase merek kacamata yang termasuk kategori kluster penjualan paling banyak (cluster_0) adalah 50%, penjualan sedang (cluster_1) adalah 27,27% dan penjualan paling sedikit (cluster_2) adalah 22,72%.

DAFTAR PUSTAKA

[1] Putra, Y.D., Sudarma, M. and Swamardika, I.B.A., 2021. Clustering History Data Penjualan Menggunakan Algoritma K-Means. *Majalah Ilmiah Teknologi Elektro, Vols*, *20*(2), pp.195-202.

[2] Sucipto, A., 2019. Klasterisasi Calon Mahasiswa Baru Menggunakan Algoritma K-Means. *Science Tech: Jurnal Ilmu Pengetahuan dan Teknologi*, *5*(2), pp.50-56.

[3] Rahayuni, S., Gunawan, I. and Kirana, I.O., 2022. Material Sales Clustering Using the K-Means Method. *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, *1*(1), pp.85-94.

[4] Marisa, F., Kom, S., Maukar, A.L., Akhriza, T.M. and MMSI, P.D., 2021. *Data Mining Konsep Dan Penerapannya*. Deepublish..

[5] Putra, R.R. and Wadisman, C., 2018. Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means. *INTECOMS: Journal of Information Technology and Computer Science*, *1*(1), pp.72-77.

[6] Mulyadien, M.K. and Enri, U., 2022. Algoritma K-Means Untuk Pengelompokan Bantuan Langsung Tunai (BLT). *Jurnal Ilmiah Wahana Pendidikan*, *8*(12), pp.198-210.

[7] Nugraha, A., Nurdiawan, O. and Dwilestari, G., 2022. Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *6*(2), pp.849-855.

[8] Qorimah, N., 2022. Implementasi Metode Naive Bayes Untuk Menentukan Minat Tabungan Pada Perumda BPR Bank Cirebon. *Jurnal Data Science & Informatika*, *2*(1), pp.15-20.

[9] Nurajizah, S., 2019. Analisa Transaksi Penjualan Obat menggunakan Algoritma Apriori. *INOVTEK Polbeng-Seri Informatika*, *4*(1), pp.35-44.

[10] Fatmawati, K. and Windarto, A.P., 2018. Data Mining: Penerapan rapidminer dengan K-means cluster pada daerah terjangkit demam berdarah dengue (DBD) berdasarkan provinsi. *CESS (Journal of Computer Engineering, System and Science)*, *3*(2), pp.173-178.

[11] Zulkifli, A., 2016. Metode C45 Untuk Mengklarifikasi Pelanggan Perusahaan Telekomunikasi Seluler. *Riau Journal Of Computer Science*, *2*(1), pp.65-76.