P-ISSN : 1858-3938
E-ISSN : 2721-284X

ALGORITMA K-MEANS CLUSTERING UNTUK MELAKUKAN SEGMENTASI PELANGGAN DI TOKO GBEE GLOW BEAUTY BUMIREJO

Siti Aisyah[1], Wachid Darmawan[2], Bambang Ismanto[3]
[1]Sistem Informasi, STMIK Widya Pratama, Kota Pekalongan
[2]Manajemen Informatika, STMIK Widya Pratama, Kota Pekalongan
[3]Teknik Informatika, STMIK Widya Pratama, Kota Pekalongan

E-mail: [1]sitiaisyahna11@gmail.com

Abstrak

Toko Gbee Glow Beauty Bumirejo merupakan toko yang menjual berbagai macam produk perawatan kulit wajah. Toko ini terletak di Jalan Pelita II No. 24 Bumirejo, Kota Pekalongan. Ada masalah dalam mengidentifikasi calon pelanggan dengan benar di Toko Gbee Glow Beauty Bumirejo. Segmentasi pelanggan dapat dilakukan menggunakan data mining dengan metode cluster untuk mengetahui potensi serta karakteristik dari setiap pelanggan. Dalam penelitian ini, metode yang digunakan adalah menggabungkan model Recency, Frekuensi, Moneter (RFM) dan algoritma k-means clustering yang digunakan untuk segmentasi pelanggan. Pembentukan cluster pelanggan didasarkan pada model RFM dan algoritma k-mean yang tepat menggunakan metode CRISP-DM. Dari penelitian yang ada di hasilkan 3 cluster pelanggan, cluster pertama berjumlah 2 pelanggan masuk dalam kategori highest loyalty, cluster kedua terdapat 13 pelanggan yang masuk dalam kategori low loyalty dan cluster ketiga berjumlah 5 pelanggan pada kategori medium loyalty. Untuk penelitian selanjutnya perlu melakukan perbandingan antara algoritma k-means dengan algoritma lainnya seperti k-medoids. Perbandingan tersebut digunakan untuk mengevaluasi hasil dari masing-masing algoritma dalam segmentasi pelanggan.
Kata kunci: Segmentasi Pelanggan, K-Means Clustering, RFM, CRISP-DM

Abstract

Gbee Glow Beauty Bumirejo Shop is a shop that sells various kinds of facial skin care products. This shop is located on Jalan Pelita II No. 24 Bumirejo, Pekalongan City. There is a problem in correctly identifying potential customers at the Gbee Glow Beauty Bumirejo Shop. Customer segmentation can be done using data mining with the cluster method to determine the potential and characteristics of each customer. In this research, the method used is to combine the Recency, Frequency, Monetary (RFM) model and the k-means clustering algorithm which is used for customer segmentation. The formation of customer clusters is based on the RFM model and the appropriate k-means algorithm using the CRISP-DM method. From the research, 3 customer clusters were produced, the first cluster had 2 customers in the highest loyalty category, the second cluster had 13 customers in the low loyalty category and the third cluster had 5 customers in the medium loyalty category. For further research, it is necessary to compare the k-means algorithm with other algorithms such as k-medoids. This comparison is used to evaluate the results of each algorithm in customer segmentation.
Key word: *Customer Segmentation, K-Means Clustering, RFM, CRISP-DM*

1. Pendahuluan

Beberapa orang menggunakan produk kecantikan untuk perawatan kulit wajah. Produk kecantikan lokal kini semakin berkembang. Dalam persaingan produk yang menarik, perusahaan harus mampu mengklasifikasikan pelanggan yang kemungkinan besar akan membeli produknya. Oleh karena itu, perusahaan harus mampu menjaring dan mengolah data pelanggan agar dapat melakukan lebih banyak penawaran (Nisa and Heikal 2022) (Herdiansyah 2015).

Toko Gbee Glow Beauty Bumirejo merupakan toko produk kecantikan dengan data transaksi yang dihasilkan setiap hari. Saat memasarkan produk kecantikan, pemilik toko memasarkannya secara online dan offline. Masalah terbesar di Toko Gbee Glow Beauty Bumirejo adalah selalu sulit untuk menentukan pelanggan potensial atau tidak, serta menentukan strategi pemasaran yang tepat untuk setiap pelanggan. Dengan adanya data transaksi pada toko tersebut memungkinkan untuk meningkatkan efisiensi dengan mengelompokkan pelanggan yang potensial atau tidak (Nurajizah 2019). Hasil kelompok ini dapat dijadikan dasar pertimbangan pengambilan

P-ISSN : 1858-3938
E-ISSN : 2721-284X

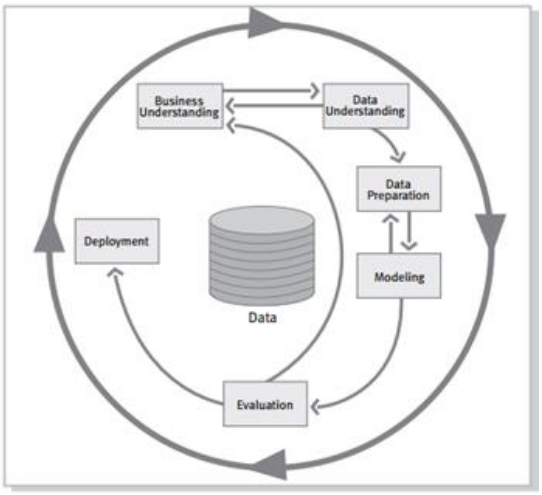
keputusan, termasuk strategi pemasaran (Ariati, et al. 2023).

Untuk mengetahui kelompok pelanggan potensial atau tidak, diperlukan sistem untuk melakukan segementasi pelanggan berdasarkan data transakasi yang ada. Sistem ini akan membantu Toko Gbee Glow Beauty Bumirejo dalam menentukan pelanggan potensial atau tidak. Penentuan pelanggan akan menggunakan gabungan analisis atau model RFM dengan algoritma k-means. Banyak penelitian sebelumnya, seperti: Pengelompokan Pelanggan di Toko Manja Beauty Skincare (Nisa and Heikal 2022), Strategi Pemasaran di PT Coversuper Indonesia Global untuk mengelompokkan pelanggan (Widiyanto and Witanti 2021), pengelompokan Pelanggan di Ova Gaming E-Sports Arena Kediri (Wijaya, Djunaidy and Mahananto 2021).

Kontribusi pada penelitian ini adalah sebagai alat bantu untuk mengetahui mengelompokkan pelanggan dan mengetahui pelanggan potinsial atau tidak. Penelitian ini penting karena di Toko Gbee Glow Beauty Bumirejo belum ada sistem yang dapat mengelompokkan pelanggan atau mencari pelanggan yang potensial atau tidak untuk menentukan kebijakan perusahaan, seperti promosi dll. Penelitian ini mengusulkan kombinasi analisis atau model RFM dengan algoritma k-mean. K-mean adalah algoritma unsupervised learning untuk mengelompokkan dataset berdasarkan transaksi yang ada (Nisa and Heikal 2022). Selain itu K-means akan digunakan untuk pembentukan klaster (Adiana, Soesanti and Permanasari 2018) dan mengelompokkan (Dinata, et al. 2020)data berdasarkan kemiripan karakteristiknya . Tujuan dari sistem yang dibuat adalah sebagai sistem untuk menetukan kelompok pelanggan yang potensial atau tidak potensial berdasarkan data transakasi yang ada di Toko Gbee Glow Beauty Bumirejo. **Referensi**.

2. Metode Penelitian

Pada penelitian ini menggunakan metode Cross Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM banyak digunakan dalam penelitian data mining (Adiana, Soesanti and Permanasari 2018). Berikut gambaran metodologi penelitian pada gambar 1.



Gambar 1. Metodologi Penelitian

2.1. Pemahaman Bisnis

Dari data transaksi yang ada di Toko Gbee Glow Beauty Bumirejo akan di terjemahkan kedalam pengetahuan data mining, dan dilanjutkan dengan menentukan stategi perusahaan kedepan kepada pelanggannya. Toko Gbee Glow Beauty Bumirejo merupakan toko kecantikan dan menjual produknya secara *offline* dan *online*. Untuk mengetahui kelompok pelanggan yang potensial atau tidak berdasarkan data transaksi yang ada menggunakan model RFM.

2.2. Pemahaman Data

Dari data transaksi yang ada di Toko Gbee Glow Beauty Bumirejo akan dilakukan pemahaman data dan dilanjutkan dengan pembersihan data awal yang tidak relevan. Penelitian ini akan menggunakan data transaksi bulan Januari-Mei 2023, di dapatkan sebanyak 145 pelanggan. Pada penelitian ini akan menggunakan sampling data sebanyak 20 pelanggan untuk dilakukan perhitungannya. Berikut sampel data penjualan di Toko Gbee Glow Beauty Bumirejo, terlihat di tabel 1

Tabel 1. Cuplikan Data Transaksi Penjualan

Kode Pelanggan	Tanggal Terakhir Transaksi	Jumlah Transaksi	Total Pengeluaran	
P1	29/01/2023	2	Rp.	100.000
P2	01/02/2023	2	Rp.	285.000
P3	26/02/2023	3	Rp.	180.000
P4	10/05/2023	2	Rp.	3.100.000
P5	01/05/2023	1	Rp.	825.000
P6	05/05/2023	2	Rp.	430.000
P7	15/05/2023	2	Rp.	80.000
P8	28/02/2023	2	Rp.	675.000

Kode Pelanggan	Tanggal Terakhir Transaksi	Jumlah Transaksi	Total Pengeluaran
P9	02/02/2023	2	Rp. 310.000
P10	30/05/2023	1	Rp. 990.000
P11	03/03/2023	1	Rp. 375.000
P12	31/01/2023	1	Rp. 495.000
P13	27/05/2023	1	Rp. 280.000
P14	27/05/2023	2	Rp. 220.000
P15	31/05/2023	2	Rp. 190.000
P16	05/05/2023	3	Rp. 655.000
P17	07/05/2023	2	Rp. 265.000
P18	25/05/2023	3	Rp. 2.715.000
P19	14/02/2023	3	Rp. 115.000
P20	02/05/2023	2	Rp. 1.225.000

2.3. Persiapan Data

Selanjutnya dari dari data transaksi yang ada di Toko Gbee Glow Beauty Bumirejo akan dilakukan *mining*. Sebelum *mining* dilakukan, data akan dipilih variabelnya, pembersihan data, dan lain sebagainya sehingga data tersebut siap untuk dilakukan transformasi data. Proses segmentasi pelanggan berdasarkan RFM dari transaksi penjualan yang ada,berdasarkan: jarak antar waktu pelanggan melakukan transaksi, jumlah pelanggan melakukan transaksi dan jumlah nominal yang di keluarkan pelanggan dalam transaksi dalam rentang waktu Januari-Mei 2023 di Toko Gbee Glow Beauty Bumirejo. Data seleksi yang dihasilkan dengan model RFM yang digunakan seperti pada tabel 2.

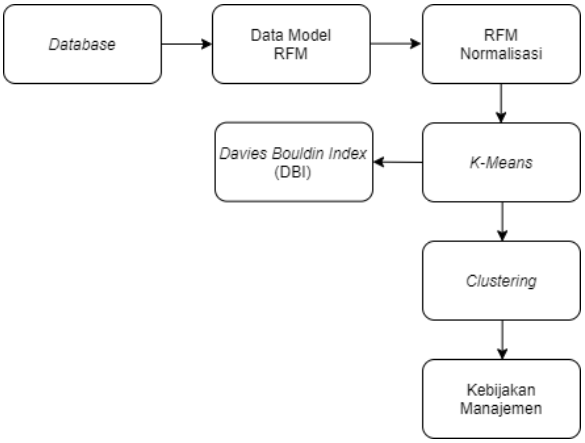
Tabel 2. Hasil Seleksi Data RFM

Data Awal	Data Akhir
Tanggal terakhir transaksi	*Recency*
Jumlah melakukan transaksi	*Frequency*
Total pengeluaran	*Monetery*

Dari seleksi data yang ada, dilanjutkan dengan pembersihan data yang tidak relevan dan dilanjutkan dengan normalisasi data, yang bertujuan agar skala data yang dihasilkan tidak terlalu jauh dengan data aslinya.

2.4. Pemodelan

Setelah dilakukan transformasi data dan proses penghapusan data yang tidak relevan, dilanjutkan dengan pemodelan menggunakan algoritma k-means. Dari database (dataset) yang sudah dikumpulkan akan dilakukan konversi data transaksi kedalam model RFM dengan melakukan pembersihan data yang tidak relevan. Setelah data siap akan dilanjutkan ketahap normalisasi data, yang bertujuan untuk menentukan skala data agar menyerupai dengan data aslinya. Selanjutnya dilakukan segementasi pelanggan menggunakan algoritma k-mean menggunakan Davies Bouldin Index (DBI) untuk menentukan nilai k yang optimal. Setelah segmentasi pelanggan terbentuk akan dilanjutkan oleh perusahaan untuk mentukan kebijakannya. Pemodelan lebih lanjut dapat dilihat pada gambar 2 berikut.



Gambar 2. Pemodelan Algoritma K-means

2.5. Evaluasi

Setelah model terbentuk dengan kualitas terbaik berdasarkan analisis data. Dilanjutkan evaluasi keefektifan dan kualitas model sebelum digunakan dan menetukan apakah sudah sesuai dengan business understanding yang ada. Evaluasi yang digunakan adalah evaluasi hasil cluster dan evaluasi cluster berdasarkan DBI.

Pada evaluasi cluster akan menggunakan 3 cluster pelanggan, seperti: *Low Loyalty*, *Medium Loyalty dan Highest Loyalty*. Sedangkan evaluasi cluster DBI digunakan untuk mengetahui nilai k yang optimal, yaitu: semakin kecil nilai DBI (*non negatif* >=0) maka cluster tersebut semakin baik.

3. Hasil Penelitian dan Pembahasan

Pada penelitian ini menggunakan 145 data transaksi secara offline yang ada di Toko Gbee Glow Beauty Bumirejo rentang waktu bulan Januari-Mei 2023. Data transaksi yang yang ada di terdiri dari kode pelanggan, tanggal transaksi, jumlah transaksi dan total transaksi. Setelah data transaksi terbentuk dilanjutkan seleksi data menggunakan model RFM, seperti pada tabel 3.

Tabel 3. Data Model RFM

Kode Pelanggan	Recency	Frequency	Monetery
P1	123	2	100000
P2	120	2	285000
P3	95	3	180000
P4	22	2	3100000

P-ISSN : 1858-3938
E-ISSN : 2721-284X

Kode Pelanggan	Recency	Frequency	Monetery
P5	31	1	825000
P6	27	2	430000
P7	17	2	80000
P8	93	2	675000
P9	119	2	310000
P10	2	1	990000
P11	90	1	375000
P12	121	1	495000
P13	5	1	280000
P14	5	2	220000
P15	1	2	190000
P16	27	3	655000
...	...	...	...
P143	42	3	265000
P144	71	3	320000
P145	75	3	690000

Setelah merubah data kedalam model RFM dilanjutkan dengan transformasi data dan normalisasi min max, agar nilai tiap atribut tidak berbeda terlalu jauh. Normalisasi min max digunakan untuk mendapatkan nilai baru dari nilai 0 sampai 1, menggunakan rumus berikut:

$$v$$

Berdasarkan perhitungan normalisasi data tersebut, didapatkan hasil normalisasi min max, seperti pada tabel 4.

Tabel 4. Data RFM Normalisasi Min Max

Kode Pelanggan	Recency	Frequency	Monetery
P1	0,8905109	0,25	0,0200617
P2	0,8686131	0,25	0,0771604
P3	0,6861313	0,5	0,0447530
P4	0,1532846	0,25	0,9459876
P5	0,2189781	0	0,2438271
P6	0,1897810	0,25	0,1219135
P7	0,1167883	0,25	0,0138889
P8	0,6715328	0,25	0,1975309
P9	0,8613138	0,25	0,0848765
P10	0,0072992	0	0,2947531
P11	0,6496350	0	0,1049383
P12	0,8759124	0	0,1419753
P13	0,0291970	0	0,0756173
P14	0,0291970	0,25	0,0570988
P15	0	0,25	0,0478395
P16	0,1897810	0,5	0,1913580
...	...	...	...
P143	0,2992700	0,5	0,0709876
P144	0,5109489	0,5	0,0879629
P145	0,5401459	0,5	0,2021604

Setelah tahap normalisasi data dilanjutkan dengan tahap pemodelan. Pada penelitian ini menggunakan algoritma k-mean untuk menentukan mengelompokkan pelanggan. Teknik clustering digunakan untuk menentukan kebijakan perusahaan terhadap stategi pemasaran yang sesuai untuk kelompong pelanggan yang sudah di tentukan. Berikut hasil algoritma k-mean terlihat pada tabel 5, menunjukkan hasil perhitungan jarak cluster.

Tabel 5. Perhitungan Jarak Cluster

	P1	P2	P3
C1	0,750684576	0,715753769	0,712644847
C2	0,278524754	0,258368418	0,5435706
C3	1,196175265	1,142507741	1,012302617

Selanjutnya dihitung ulang jarak *cluster* sampai posisi data tidak mengalami perubahan (*iterasi* 1) terlihat pada tabel 6 dan *iterasi* 2 pada tabel 7.

Tabel 6. Perhitungan Jarak *Cluster* (*Iterasi 1*)

	P1	P2	P3
C1	0,795921552	0,767379959	0,674056386
C2	0,134610689	0,095399133	0,268247827
C3	1,005328163	0,953195701	0,830416617

Tabel 7. Perhitungan Jarak *Cluster* (*Iterasi 2*)

	P1	P2	P3
C1	0,785895665	0,756851186	0,65184813
C2	0,125229819	0,084516145	0,27201991
C3	1,180544958	1,124183151	1,034191008

Dari perhitungan k-mean yang sudah dilakukan, maka hasil tiap *cluster* pelanggan terlihat pada tabel 8.

Tabel 8. Hasil *Clustering*

Kode Pelanggan	Recency	Frequency	Monetery	Cluster
P1	0,8905109	0,25	0,0200617	C2
P2	0,8686131	0,25	0,0771604	C2
P3	0,6861313	0,5	0,0447530	C2
P4	0,1532846	0,25	0,9459876	C1
P5	0,2189781	0	0,2438271	C3
P6	0,1897810	0,25	0,1219135	C2
P7	0,1167883	0,25	0,0138888	C2
P8	0,6715328	0,25	0,1975308	C3
P9	0,8613138	0,25	0,0848765	C2
P10	0,0072992	0	0,2947530	C2
P11	0,6496350	0	0,1049382	C2
P12	0,8759124	0	0,1419753	C2
P13	0,0291970	0	0,0756172	C2
P14	0,0291970	0,25	0,0570987	C2
P15	0	0,25	0,0478395	C2
P16	0,1897810	0,5	0,1913580	C3
...	...	...	...	...
P143	0,2992700	0,5	0,0709876	C2
P144	0,5109489	0,5	0,0879629	C2
P145	0,5401459	0,5	0,2021604	C3

P-ISSN : 1858-3938
E-ISSN : 2721-284X

Jadi dapat disimpulkan bahwa sebaran cluster sebagai berikut: C1 sebanyak 2 pelanggan, C2: sebanyak 13 pelanggan dan C3: sebanyak 5 pelanggan. Untuk menentukan apakah nilai k yang digunakan sudah optimal atau belum pada algoritma k-mean menggunakan DBI. Dari perhitungan yang sudah dilakukan nilai DBInya sebesar 0,661142081. Berdasarkan data tersebut maka nilai k yang digunakan dalam menentukan *cluster* pada algoritma k-mean dikatakan baik, karena nilai yang didapatkan *non negatif* >=0 (mendekati nilai 0).

Dari perhitungan algoritma k-mean dilanjutkan dengan melakukan analisis hasil k-mean untuk menentukan stategi perusahaan untuk masing-masing pelanggannya, seperti terlihat di tabel 9.

Tabel 9. Simbol *Rank* Setiap *Cluster*

Cluster	Jumlah Pelanggan	Simbol Rank
C1	22	R ↓ F ↑ M ↑
C2	86	R ↓ F ↓ M ↓
C3	37	R ↑ F ↓ M ↓

Dari tabel 9 dapat di jelaska, bahwa:

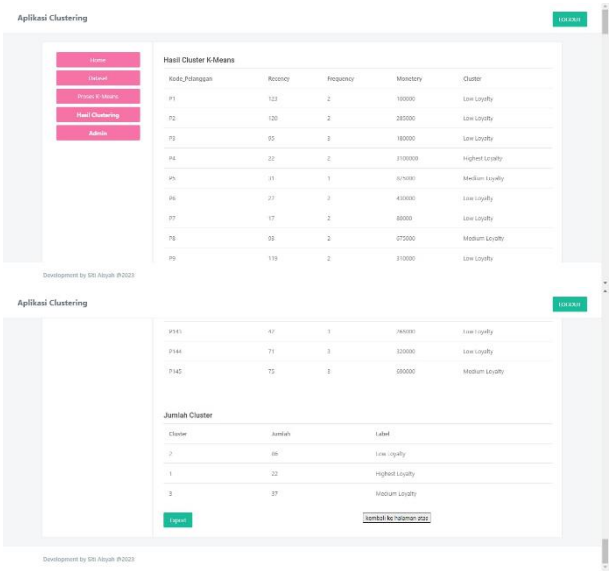
a. *Cluster* 1, nilai rata-rata *recency:* 0,098540146, berarti memiliki nilai *recency* yang dekat. Untuk nilai rata-rata *frequency:* 0,375, berarti memiliki nilai *frequency* yang tinggi. Sedangkan nilai rata-rata *monetery:* 0,886574, yang berarti memiliki nilai *monetery* tinggi.
b. *Cluster* 2, nilai rata-rata *recency:* 0,284671533, berarti memiliki nilai *recency* yang sedang. Untuk nilia rata-rata *frequency* : 0,21, sehingga memiliki nilai *frequency* yang sedang. Sedangkan *monetery* nilai rata-rata: 0,086998, yang berarti memiliki nilai *monetery* rendah.
c. *Cluster* 3, pada *recency* memiliki nilai rata-rata: 0,516788321, yang berarti memiliki nilai *recency* nya jauh. Untuk nilai *frequency* rata-rata: 0,2, sehingga memiliki nilai *frequency* yang rendah. Sedangkan nilai *monetery* rata-rata: 0,148457, yang berarti memiliki nilai *monetery* sedang.

Berdasarkan simbol *rank* yang dihasilkan dapat disimpulkan bahwa kelompok C1: masuk kedalam pelanggan jenis *highest loyalty*, kelompok C2: pelanggan jenis *low loyalty* dan kelompok C3: termasuk pelanggan jenis *medium loyalty*, seperti pada tabel 10.

Tabel 10. Hasil *Clustering*

Cluster	Jumlah Pelanggan	label
C1	22	Highest Loyalty
C2	86	Low Loyalty
C3	37	Medium Loyalty

Berikut hasil cluster dan labelnya dari system yang sudah dibuat berdasarkan perhitungan model RFM dan algoritma k-mean dengan jumlah data transaksi sebanyak 145, seperti pada gambar 3.



Gambar 3. Hasil *Clustering* di Aplikasi

4. Kesimpulan

Untuk mempermudah pengelompokan jenis pelanggan di Toko Gbee Glow Beauty Bumirejo perlu adanya sistem yang bisa menanganinya dengan mudah dan cepat. Penelitian ini mengusulkan segmentasi pelanggan dengan kombinasi model RFM dan algoritma k-mean untuk membantu Toko Gbee Glow Beauty Bumirejo dalam pengelompokan pelanggan agar perusahaan bisa dengan cepat dalam menangani pelanggan yang loyal maupun tidak loyal. Berdasarkan aplikasi yang dibuat di dapatkan 3 jenis pelanggan, diantaranya: pelanggan jenis highest loyalty ada di cluster 1, cluster 2: pelanggan jenis low loyalty dan cluster 3: termasuk pelanggan jenis medium loyalty.

Untuk penelitian kedepan bisa dilakukan perbandingan algoritma untuk mengevaluasi algoritma yang digunakan saat ini, misalnya membandingkan dengan algoritma k-medoids. Selain itu perlu adanya konversi data transaksi secara otomatis ke dalam bentuk data RFM.

5. Referensi

Adiana, Beta Estri, Indah Soesanti, and Adhistya Erna Permanasari. 2018. "Analisis Segmentasi Pelanggan Menggunakan Kombinasi Rfm Model Dan Teknik Clustering." *JUTEI: Jurnal Terapan Teknologi Informasi* 23-32.

Ariati, Ira, Reza Nugraha Norsa, Lurinjani Akhsan, and Jerry Heikal. 2023. "Segmentasi Pelanggan Menggunakan K-means

P-ISSN : 1858-3938
E-ISSN : 2721-284X

Clustering, Studi Kasus: Pelanggan UHT Milk Greenfield." *Cerdika: Jurnal Ilmiah Indonesia* 629-643.

Dinata, Rozzi Kesuma, Haried Novriando, Novia Hasdyna, and Sujacka Retno. 2020. "Reduksi Atribut Menggunakan Information Gain untuk Optimasi Cluster Algoritma K-Means." *JEPIN: Jurnal Edukasi dan Penelitian Informatika* 48-53.

Herdiansyah, Haris. 2015. *Wawancara, Observasi, Dan Focus Groups: Sebagai Instrumen Penggalian Data Kualitatif.* Jakarta: Rajawali pers.

Nisa, Khairi, and Jerry Heikal. 2022. "Strategi Segmentasi Pelanggan Manja Beauty Skincare Dengan Menggunakan Analisa RFM Model." *JATI (Jurnal Mahasiswa Teknik Informatika)* 348-351.

Nurajizah, Siti. 2019. "Analisa Transaksi Penjualan Obat menggunakan Algoritma Apriori." *Jurnal Inovtek Polbeng Seri Informatika* 35-44.

Raghuwanshi, Shailendra Singh, and Prem Narayan Arya. 2012. "Comparison of K-means and Modified K-mean algorithms for Large Data-set." *International Journal of Computing, Communications and Networking* 106-110.

Singh, Ran Vijay, and M. P. S Bhatia. 2011. "Data clustering with modified K-means algorithm." *International Conference on Recent Trends in Information Technology*. Chennai: ICRTIT. 717-721.

Widiyanto, Agus Tri, and Arita Witanti. 2021. "Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global)." *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi* 204-215.

Wijaya, Kartika Zahretta, Arif Djunaidy, and Faizal Mahananto. 2021. "Segmentasi Pelanggan Menggunakan Algoritma K-Means dan Analisis RFM di Ova Gaming E-Sports Arena Kediri." *Jurnal TEKNIK ITS* 230-237.