

ANALISA TINGKAT PENJUALAN PRODUK MENGGUNAKAN K-NEAREST NEIGHBOR (K-NN) DAN K-MEANS (STUDI KASUS PERUSAHAAN KAYU ELANG PERKASA)

Herianto¹, Nur Syamsiyah², Sapitri Anggraini³

^{1,3}Dosen Program Studi Teknologi Informasi Universitas Darma Persada

²Program Studi Sistem Informasi Universitas Darma Persada

email : heri.unsada@gmail.com

ABSTRAK

Tujuan dari penelitian ini adalah memperoleh pengetahuan tingkat penjualan produk pada sebuah perusahaan untuk memberikan rekomendasi pada pemilik perusahaan tentang produk mana yang prioritas diberi perhatian demi kemajuan perusahaan tersebut. Metodologi pada penelitian ini menggunakan metodologi CRISP-DM (CRoss Industry Standard Process for Data Mining). Pada penelitian ini algoritma yang digunakan adalah jenis klasifikasi dan clustering yaitu K-NN dan K-Means. Dari penelitian ini akan diuji yang lebih baik apakah menggunakan analisa klasifikasi atau clustering dengan menggunakan data yang sama. Berdasarkan hasil pengujian akurasi kedua metode diperoleh bahwa K-Means memiliki tingkat akurasi yang lebih baik sebesar 78,37% sedangkan K-NN memiliki tingkat akurasi 76,06%.

Kata kunci : Algoritma K-Means, Clustering, Algoritma K-NN, Classification, Data Mining, PHP, MySQL

1. PENDAHULUAN

Pk Elang Perkasa adalah sebuah perusahaan milik pribadi yang bergerak di bidang penjualan kayu yang menjual berbagai macam jenis kayu dengan berbagai macam ukuran. Pada kegiatannya Pk. Elang Perkasa melakukan proses jual beli yang cukup banyak dengan berbagai jenis kayu dan macam ukurannya. Dengan banyaknya transaksi penjualan sehari-hari maka data penjualan pun semakin lama akan semakin bertambah banyak.

Penjualan yang semakin tinggi tersebut membutuhkan informasi dan identifikasi produk mana yang paling potensial dan produk mana yang kurang potensial pada penjual. Dengan adanya pengelompokan pihak perusahaan dapat mengetahui barang yang sangat laris, cukup laris dan kurang laris. Sehingga barang yang ada digudang tidak menumpuk dan sesuai dengan kebutuhan. Dengan adanya pengolahan data yang dilakukan diharapkan akan dapat memberikan solusi nyata kepada pihak perusahaan. dan diharapkan pemilik perusahaan dapat melakukan pemasaran dengan strategi yang tepat untuk melayani kebutuhan konsumen.

Secara komputerisasi permasalahan ini bisa dibantu oleh metode atau algoritma yang dapat mengelompokkan dan menganalisis hasil transaksi penjualan di atas. Penggunaan penerapan pada metode yang berbeda dengan suatu data yang sama dapat menghasilkan suatu nilai keakuratan yang berbeda pula. Penggunaan metode yang berbeda dapat dibandingkan untuk mengetahui kelebihan ataupun keakuratan dalam mengklasifikasi masing-masing metode dengan menggunakan data yang sama.

1.1. Data Mining

Menurut (Efori Buulolo, 2020) dalam *e-book* yang berjudul “*Data mining* untuk Perguruan Tinggi”. *Data Mining* atau kadang disebut juga *Knowledge Discovery in Database* (KDD) merupakan aktivitas yang berkaitan dengan pengumpulan data, pemakaian data historis untuk menemukan pengetahuan, informasi, keteraturan, pola atau hubungan dalam data yang berukuran besar. *Output* dalam *data mining* dapat dipergunakan sebagai alternatif dalam pengambilan keputusan atau untuk memperbaiki keputusan di masa yang akan datang.

1.2. Metode K-Means Clustering

Algoritma *K-Means* merupakan metode non-hierarki yang pada awalnya mengambil sebagian banyaknya komponen populasi untuk dijadikan pusat *cluster* awal. Pada tahap ini pusat *cluster* dipilih secara acak dari sekumpulan populasi data. Berikutnya *K-Means* menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat *cluster* yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap kluster. Posisi pusat kluster akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap-tiap pusat kluster dan terakhir akan terbentuk posisi pusat *Muster* yang baru (Danni dan Setiawan, 2016).

Langkah-langkah pengelompokan data adalah :

1. Pilih jumlah *Master*.
2. Inisialisasi awal dan pusat kluster dilakukan secara random.
3. Setiap data ditempatkan ke pusat kluster terdekat berdasarkan jarak antar obyek. Pada tahap ini jarak dihitung dengan menentukan kemiripan atau ketidakmiripan data dengan Metode Jarak Euclidean (Euclidean Distance) dengan rumus seperti dibawah ini (Witanto, Ratnawati dan Anam, 2019) :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

dimana :

$d(x, y)$ = ukuran ketidakmiripan

$x_i = (x_1, x_2, \dots, x_i)$ yaitu variabel data

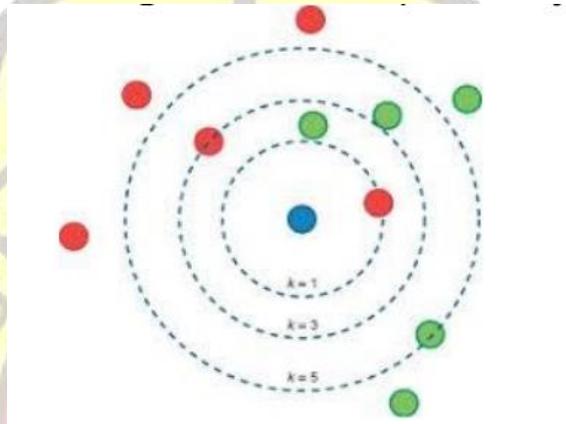
$y_i = (y_1, y_2, \dots, y_j)$ yaitu variabel pada titik pusat.

4. Hitung Pusat kluster yang baru dengan cara menghitung rata-rata obyek pada kluster. Penghitungan bisa juga dengan menggunakan median.

5. Hitung Kembali jarak tiap objek dengan pusat kluster yang baru, hingga kluster tidak berubah, maka proses pengklasteran selesai.

1.3. Metode K-NN

Metode *K-Nearest Neighbor* (K-NN) merupakan salah satu dalam top 10 metode data mining yang paling banyak digunakan. Metode ini melakukan klasifikasi berdasarkan kemiripan suatu data dengan data yang lain. Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan training sample (Herianto Herianto (Universitas Darma Persada et al., 2020). Prinsip sederhana metode ini adalah “Jika suatu hewan berjalan seperti bebek, bersuara kwek-kwek seperti bebek, dan penampilannya seperti bebek, maka hewan itu mungkin bebek” (Prasetyo, 2014).



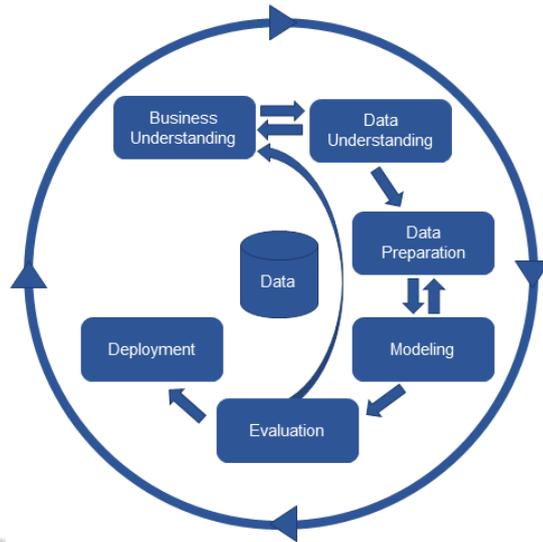
Gambar 1. Ilustrasi K-NN

Langkah-langkah klasifikasi data menggunakan K-Nearest Neighbor adalah sebagai berikut (Ibrahim, Bacheramsyah, & Hidayat, 2018):

1. Tentukan nilai K
2. Hitung jarak antara data baru ke setiap label data
3. Tentukan k labeled data yang mempunyai jarak yang paling minimal
4. Klasifikasikan data baru ke dalam label data yang mayoritas K-NN dipilih berdasarkan metrik jarak.

2. METODOLOGI

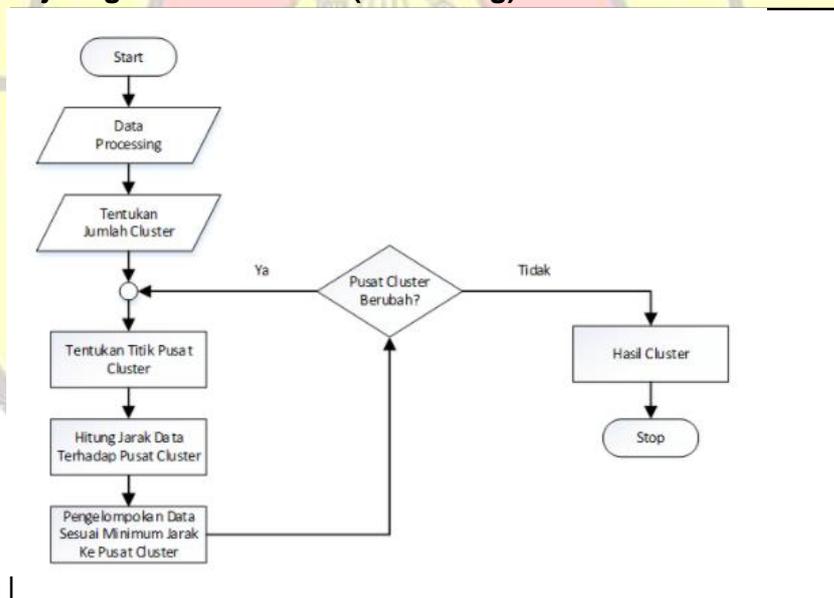
Penelitian ini di bidang data science atau data mining menggunakan tahapan CRISP-DM (*CRoss Industry Standard Process for Data Mining*) seperti berikut :



Gambar 2. Metodologi CRISP-DM

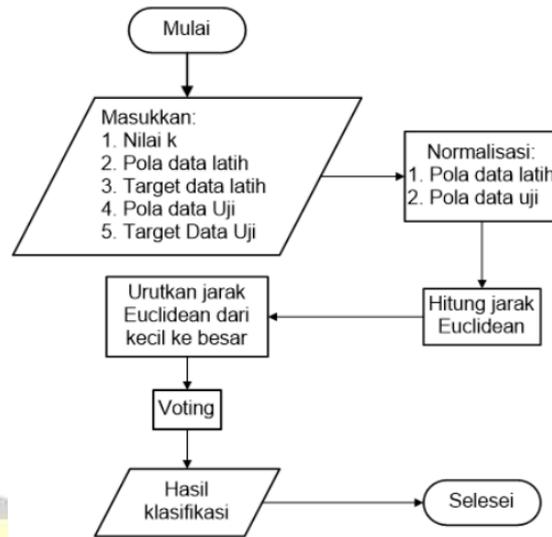
Algoritma data mining yang digunakan adalah metoda K-NN dan K-Mean dengan alur kerja berikut :

2.1. Alur Kerja Algoritma K-Means (Clustering)



Gambar 3. Flowchart Algoritma K-Means Clustering

2.2. Alur Kerja Algoritma K- Nearest Neighbor (Klasifikasi)



Gambar 4. Flowchart Algoritma K-NN

3. HASIL DAN PEMBAHASAN

3.1. Perbandingan Metode K-Means dan K-Nearest Neighbor Algoritma K-Means Clustering

a) Data Training

Data training ini dikelompokkan menjadi 3 cluster yaitu cluster 1, cluster 2 dan cluster 3. Dimana data *training* akan diolah dengan algoritma *K-Means Clustering* serta mengeluarkan hasil nilai prediksi analisis setiap data yang telah di *training*.

Tabel 1. Sample Data Training

Volume	Harga	Jenis Barang	Jumlah Order
0,21	1	1	4
0,064	3	1	2
0,144	2	1	3
0,0576	3	2	2
0,0408	2	2	1
0,048	1	3	1

b) Inisialisasi Data

Dilakukan pemilihan K data sebagai centroid awal, misalnya dipilih data ke 1, 5 dan 16. Inisialisasi *centroid* dapat ditentukan secara manual ataupun random.

Tabel 2. Nilai Buah (K)

Centroid	Volume	Harga	Jenis Barang	Jumlah Order
C1	0,21	1	1	4
C2	0,0408	2	2	1
C3	0,048	1	3	1

c) Iterasi 1

Menghitung jarak setiap data ke *centroid* data terdekat. *Centroid* terdekat akan menjadi *cluster* yang diikuti oleh data tersebut. Berikut perhitungan jarak ke setiap *centroid* pada data ke-i.

$$\begin{aligned} d(i - C_{1i}) &= \sqrt{(a_1 - C_{1a})^2 + (b_1 - C_{1b})^2 + (c_1 - C_{1c})^2 + (d_1 - C_{1d})^2} \\ &= \sqrt{(0,21 - 0,21)^2 + (1 - 1)^2 + (1 - 1)^2 + (4 - 4)^2} = 0 \end{aligned} \quad (2)$$

.....

$$\begin{aligned} &= \sqrt{(0,048 - 0,048)^2 + (2 - 1)^2 + (2 - 3)^2 + (1 - 1)^2} \\ &= 2,0072 \\ &= \sqrt{(0,048 - 0,048)^2 + (1 - 1)^2 + (3 - 3)^2 + (1 - 1)^2} = 0 \end{aligned}$$

Rangkuman hasilnya seperti pada table berikut :

Tabel 3. Data Hasil Perhitungan Iterasi 1

Data ke-i	Jarak ke centroid			Cluster yang diikuti
	1	2	3	
1	0	9,1524	13,162	c1
2	8,146	1,0064	9,016	c2
3	2,066	3,0864	9,096	c1
4	9,1524	0	6,0096	c2
5	11,1692	2,0168	2,0072	c3
6	13,162	6,0096	0	c3

Selanjutnya dihitung *centroid* yang baru untuk setiap *cluster* berdasarkan data yang bergabung pada setiap *cluster* nya. Untuk *cluster* 1, ada 2 data yang bergabung ke dalamnya, *cluster* 2 ada 2 data yang bergabung ke dalamnya dan *cluster* 3 ada 2 data yang bergabung ke dalamnya.

d) Iterasi 2

Menghitung jarak setiap data ke *centroid* terdekat. *Centroid* terdekat akan menjadi *cluster* yang diikuti oleh data tersebut. Berikut perhitungan jarak ke setiap *centroid* pada data ke-l menggunakan *centroid* yang baru.

Tabel 4. Data Centroid Baru

Centroid	A	B	C	d
C1	0,177	1,5	1	3,5
C2	0,0608	3	1,5	2
C3	0,0444	1,5	2,5	1

$$\begin{aligned} d(i - C_{1i}) &= \sqrt{(a_1 - C_{1a})^2 + (b_1 - C_{1b})^2 + (c_1 - C_{1c})^2 + (d_1 - C_{1d})^2} \\ &= \sqrt{(0,21 - 0,177)^2 + (1 - 1,5)^2 + (1 - 1)^2 + (4 - 3,5)^2} \\ &= 0,533 \end{aligned} \quad (3)$$

$$= \sqrt{(0,048 - 0,0444)^2 + (1 - 1,5)^2 + (3 - 2,5)^2 + (1 - 1)^2}$$

$$= 0,5036$$

Rangkuman proses di atas seperti pada tabel berikut :

Tabel 5. Data Hasil Perhitungan Iterasi 2

Data ke-i	Jarak ke centroid			Cluster yang diikuti
	1	2	3	
1	0,533	8,3992	11,6656	C1
2	4,613	0,2532	5,5196	C2
3	0,533	2,3332	6,5996	C1
4	5,6194	0,2532	3,5132	C2
5	7,6362	2,27	0,5036	C3
6	10,629	7,2628	0,5036	C3

Dalam metode/algorithm *K-Means Clustering*, pengelompokan suatu *cluster* yang dihitung dengan *centroid* yang baru dilakukan secara berulang sampai tidak ada data yang berubah. Karena tidak adanya nilai cluster yang berubah lagi, maka proses iterasi selesai.

3.2. Metode K-Nearest Neighbor

Langkah awal dalam *system* ini adalah menyiapkan data set atau data training penjualan kayu yang diberikan label sangat laris, cukup laris dan kurang laris. Data set ini akan digunakan untuk acuan dalam mentraining data yang akan diolah menggunakan metode *K-Nearest Neighbor*.

Tabel 6. Cuplikan Data Set

Nama Barang	Volume	Harga	Jenis Barang	Jumlah Order	Label
Merbau Oven 500 x 3 x 35	0,21	1	1	4	Kurang Laris
Samarinda Oven 400 x 4 x 20	0,064	3	1	2	Sangat Laris
Samarinda Oven 400 x 6 x 17	0,0408	2	2	1	Cukup Laris
Merbau Oven 400 x 8 x 15	0,048	1	3	1	Kurang Laris

Tabel 7. Data Testing

Nama Barang	Volume	Harga	Jenis Barang	Jumlah Order	Label
Merbau Oven 400 x 5 x 15	0,03	2	2	1	?

1. Menentukan nilai k (tetangga terdekat) Misalkan K=3.

2. Melakukan Normalisasi pada data *training*.

Min-max normalization:

$$X^* = \frac{X - \min(X)}{\text{range}(X)} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (4)$$

Normalisasi Volume

1. $(0,21 - 0,0408) / (0,21 - 0,0408) = 1$
2. $(0,064 - 0,0408) / (0,21 - 0,0408) = 0,137115839$
3. $(0,0408 - 0,0408) / (0,21 - 0,0408) = 0$
4. $(0,048 - 0,0408) / (0,21 - 0,0408) = 0,042553191$

Normalisasi Harga

1. $(1 - 1) / (3 - 1) = 0$
2. $(3 - 1) / (3 - 1) = 1$
3. $(2 - 1) / (3 - 1) = 0,5$
4. $(1 - 1) / (3 - 1) = 0$

Normalisasi Jenis Barang

1. $(1 - 1) / (3 - 1) = 0$
2. $(1 - 1) / (3 - 1) = 0$
3. $(2 - 1) / (3 - 1) = 0,5$
4. $(3 - 1) / (3 - 1) = 1$

Normalisasi Jumlah Order

1. $(4 - 1) / (4 - 1) = 1$
2. $(2 - 1) / (4 - 1) = 0,333333333$
3. $(1 - 1) / (4 - 1) = 0$
4. $(1 - 1) / (4 - 1) = 0$

Tabel 8. Hasil Normalisasi

Volume	Harga	Jenis Barang	Jumlah Order
1	0	0	1
0,137115839	1	0	0,333333333
0	0,5	0,5	0
0,042553191	0	1	0

3. Menghitung jarak antara data uji dengan data *training* dengan *Euclidean Distance*.Hitung jarak data uji dengan data *training* pertama

$$\text{Jarak} = \sqrt{(0,03 - 1)^2 + (2 - 0)^2 + (2 - 0)^2 + (1 - 1)^2} = 8,97$$

$$\text{Jarak} = \sqrt{(0,03 - 0,137115839)^2 + (2 - 1)^2 + (2 - 0)^2 + (1 - 0,333333333)^2} = 5,551560284$$

$$\text{Jarak} = \sqrt{(0,03 - 0)^2 + (2 - 0,5)^2 + (2 - 0,5)^2 + (1 - 0)^2} = 5,53$$

$$\text{Jarak} = \sqrt{(0,03 - 0,042553191)^2 + (2 - 0)^2 + (2 - 1)^2 + (1 - 0)^2} = 6,012553191$$

Gambar 9. Hasil Perhitungan Jarak

Volume	Harga	Jenis Barang	Jumlah Order	Euclidean Distance	Rangking	Label
Merbau Oven 500 x 3 x 35	0,21	1	1	8,97	1	Kurang Laris
Samarinda Oven 400 x 4 x 20	0,064	3	1	5,551560284	3	Sangat Laris
Samarinda Oven 400 x 6 x 17	0,0408	2	2	5,53	4	Cukup Laris
Merbau Oven 400 x 8 x 15	0,048	1	3	6,012553191	2	Kurang Laris

4. Setelah menghitung jarak satu per satu selanjutnya mencari jarak yang terkecil dengan mengurutkan hasil perhitungan jarak secara *ascending* (berurutan dari jarak terkecil ke jarak terbesar).

Gambar 10. Urut Hasil Perhitungan Jarak

Volume	Harga	Jenis Barang	Jumlah Order	Euclidean Distance	Rangking	Label
Merbau Oven 500 x 3 x 35	0,21	1	1	8,97	1	Kurang Laris
Merbau Oven 400 x 8 x 15	0,048	1	3	6,012553191	2	Kurang Laris
Samarinda Oven 400 x 4 x 20	0,064	3	1	5,551560284	3	Sangat Laris
Samarinda Oven 400 x 6 x 17	0,0408	2	2	5,53	4	Cukup Laris

5. Selanjutnya mengumpulkan jarak terkecil yang telah diurutkan berdasarkan inputan nilai k (tetangga terdekat). Karena nilai k = 3, maka yang diambil 3 terkecil.

Gambar 11. Tetangga Terdekat K=3

Volume	Harga	Jenis Barang	Jumlah Order	Euclidean Distance	Rangking	Label
Merbau Oven 500 x 3 x 35	0,21	1	1	8,97	1	Kurang Laris
Merbau Oven 400 x 8 x 15	0,048	1	3	6,012553191	2	Kurang Laris
Samarinda Oven 400 x 4 x 20	0,064	3	1	5,551560284	3	Sangat Laris

5. Berdasarkan table 4.11 diatas label yang banyak muncul adalah kurang laris, maka hasilnya adalah kurang laris.

3.3. Pengukuran dan Evaluasi

1. Akurasi K-Means = $\frac{\text{Prediksi yang benar}}{\text{total data yang diprediksi}} \times 100 = \frac{203}{259} \times 100 = 78,37\%$
2. Akurasi K-NN = $\frac{\text{Prediksi yang benar}}{\text{total data yang diprediksi}} \times 100 = \frac{197}{259} \times 100 = 76,06\%$

Akurasi metode K-Means 78,37% sedangkan akurasi K-NN yaitu 76,06% maka metode K-Means lebih tinggi akurasinya dibandingkan metode K-NN.

4. KESIMPULAN

1. Penelitian ini berhasil membangun dan merancang sebuah analisis penjualan yang dapat digunakan untuk membantu pihak perusahaan dalam menyusun strategi penjualan agar memenuhi kepuasan pelanggan.
2. Berdasarkan hasil pengujian dengan mengukur kinerja kedua metode tersebut menggunakan rumus akurasi, diperoleh bahwa K-Means memiliki tingkat akurasi yang lebih baik yakni sebesar 78,37% sedangkan K-NN memiliki tingkat akurasi yakni 76,06%.

DAFTAR PUSTAKA

1. Bahar, Apriadi, Bambang Pramono, Laode Hassanuddin S Sagala, 2016, ***Penentuan Strategi Penjualan Alat-Alat Tatto Di Studio Sonyxtatto Menggunakan Metode K-Means Clustering***, ISSN: 2502-8928. Vol.2 (hlm 75-86).
2. Herianto Herianto, Cahyaningrum, Nila, 2020, ***Implementasi K-Nn Dan Ahp Untuk Rekomendasi Model Pakaian Toko Online***, *Jurnal Sains & Teknologi Fakultas Teknik, X (2)*. Pp. 9-19. ISSN 2088-060X, X(1), 9–19. Retrieved from <http://repository.unsada.ac.id/1631/>
3. Meliala, Meilida Dina, Penda Hasugian, 2020, ***Perbandingan Algoritma K-Nearest Neighbor Dengan Decision Tree Dalam Memprediksi Penjualan Makanan Hewan Peliharaan Di Petshop Dore Vet Clinic***, *Jurnal Teknologi Informasi*. ISSN: 1907-2430. Vol. XV (hlm 35-39).
4. Pradnyana, Aditra, Gede, Agus Aan Jiwa Permana, 2017, ***Perbandingan Algoritma K-Means Dan Hybrid K-Means KNN Untuk Pembagian Kelas Kuliah Mahasiswa***, Seminar Nasional Riset Inovatif. ISBN: 978-602-6428-11-0 (hlm 941-949).
5. Prayoga, Yudi, Heru Satria Tambunan, Iin Parlina, 2019, ***Penerapan Clustering Pada Laju Inflasi Kota di Indonesia Dengan Algoritma K-Means***, *Jurnal Penerapan Kecerdasan Buatan*. Vol. 1, No. 1(hlm 24-30).
6. Sembiring, Falentino, Octavia dan Sudin Saepudin, 2020, ***Implementasi Metode K-Means Dalam Pengklasteran Daerah Pungutan Liar Di Kabupaten Sukabumi*** (Studi Kasus : Dinas Kependudukan Dan Pencatatan Sipil). Vol. 14, No.1 (hlm 40-47).
7. Suyanto, 2017, ***Data Mining untuk Klasifikasi dan Klasterisasi Data***, Bandung: Informatika Bandung.
8. Umam, Khaerul dan Muhammad Hilman Fakhriza, 2021, ***Analisis Produk Terlaris Menggunakan Metode K-Means Clustering Pada PT. Sukanda Jaya***, *Jurnal Informatika*. E-ISSN: 2722-2713. Hlm. 8-15.

9. Utomo, Putro, Dito dan Mesran, 2020, ***Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung***, Jurnal Media Informatika Budidarma. ISBN: 2614-5278. ISSN : 2548-8368 (hlm 437-444).
10. Vlandari, R.T, 2017, ***Data Mining Teori dan Aplikasi Rapidminer***, Yogyakarta: Gava Media.
11. Widaningsih, Sri, 2019, ***Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, Naïve Bayes, K-NN Dan SVM***, Jurnal Tekno Insentif. ISSN(p):1907-4964. Vol.13, No 1 (hlm 16-25).
12. Yolanda, Ike dan Hasanul Fahmi, 2021, ***Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT. Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor***, Jurnal Ilmu Komputer dan Sistem Informasi. E-ISSN: 2723-6129. Hlm 9-15.

