

PENERAPAN DATA MINING UNTUK MENENTUKAN JENIS ASURANSI YANG PALING DIMINATI DI PT AIA FINANCIAL DENGAN METODE CLUSTERING

Fahrul Razi¹, Vany Terisia², Diana Yusuf³ (*)

¹ITB Ahmad Dahlan, Jakarta

²ITB Ahmad Dahlan, Jakarta

³ITB Ahmad Dahlan, Jakarta

Abstract

Finding information that is concealed in databases is referred to as data mining. Data mining is a process used to extract and find relevant information and related knowledge from numerous very big datasets using statistical, mathematical, artificial intelligence, and machine learning approaches. The steps involved in the data mining process start with the selection of data from the source data to the target data, followed by preprocessing to enhance the quality of the data, transformation, data mining, and stages of interpretation and assessment that result in new information. Organizing data sets into many groups through the process of clustering ensures that objects in one group share a lot in common and differ greatly from those in other groups. Data mining technique K-Means Clustering divides data into one or more clusters. RapidMiner is the program used in this data mining application, which uses the Knowledge Discovery in Databases (KDD) stage. According to this report, AIA Family First Protection, AIA Prolink, and AIA Life Secure are the three most popular insurance products. With the use of this research, PT. AIA Financial will be able to identify the most popular insurance product categories and market these products to the general public.

Kata Kunci: Assurance, Data Mining, K-Means Clustering, RapidMiner

Januari – Juni 2023, Vol 4 (1) : hlm 1-13
©2023 Institut Teknologi dan Bisnis Ahmad Dahlan.
All rights reserved.

(*) Korespondensi: fahrulrazio398@gmail.com (Fahrul Razi), vterisia@gmail.com (Vany Terisia), dianayusuf01@gmail.com (Diana Yusuf)

PENDAHULUAN

Dalam era globalisasi, perkembangan kecanggihan teknologi yang semakin pesat merupakan aspek yang dapat dimanfaatkan untuk mencapai kemudahan-kemudahan, tidak terkecuali dalam arus informasi (Metisen dan Sari, 2015). Kemajuan teknologi informasi sudah semakin berkembang pesat dalam segala bidang kehidupan. Banyak sekali data yang dihasilkan oleh teknologi informasi yang canggih, mulai dari bidang ekonomi, industri, dan teknologi serta berbagai bidang kehidupan lainnya (Yunita, 2018). Seharusnya data yang telah tersedia dapat dijadikan sebagai sistem pengambilan keputusan untuk solusi bisnis serta dukungan infrastruktur di bidang teknologi yang merupakan penyebab munculnya suatu teknologi yaitu data mining (Normah, Nurajizah dan Salbinda, 2021).

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan (Hendrian, 2018). Data mining adalah proses menemukan pola dan pengetahuan menarik dari data dalam jumlah yang besar (Nabila, et al., 2021). Data mining yang dimaksud untuk memberikan solusi nyata bagi para pengambil keputusan di dunia business untuk mengembangkan bisnis mereka (Sutrisno, Afriyudi dan Widiyanto, 2013). Data mining, sering juga disebut sebagai knowledge discovery in database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Tamba, Kesuma dan Feryanto, 2019). Dengan adanya Data mining, banyak perusahaan yang telah menggunakan metode ini untuk kegiatan bisnis dalam perusahaan. Sebagai contoh, sebuah perusahaan Asuransi AIA Financial yang membutuhkan data mining untuk menentukan jenis asuransi yang paling banyak diminati oleh pengguna. Dengan mengetahui jenis asuransi yang diminati, maka perusahaan dapat memprediksi pasar dan mendapatkan informasi baru dari hasil data mining yang didapat.

Analisis Pengelompokan / Clustering merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar daripada kesamaan data tersebut dengan data dalam kelompok lain. Potensi clustering adalah dapat digunakan untuk mengetahui struktur dalam data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luas seperti klasifikasi, pengolahan gambar, dan pengenalan pola (Sadewo, Windarto dan Hartama, 2017). Ada banyak metode yang tersedia untuk memperkirakan jumlah cluster antara lain indeks statistik, metode berbasis varians, Teori Informasi, dan metode goodness of fit. Penelitian ini mengeksplorasi enam pendekatan yang berbeda untuk menentukan jumlah cluster yang tepat dalam suatu dataset (Handoko, Fauziah dan Handayani, 2020).

Data hasil penjualan produk asuransi yang digunakan adalah data hasil perusahaan dalam satu tahun terakhir. Terdapat beberapa kriteria yang digunakan, yaitu Data jenis-jenis produk asuransi, Data banyak nasabah yang mengambil suatu produk. Dengan menggunakan metode Clustering, perusahaan mendapatkan informasi bisnis yang baru dari hasil penelitian ini. Dan manfaat dari penelitian ini adalah menemukan peluang bisnis baru yang didapat dari pengolahan Database, Mengetahui jenis asuransi mana yang paling diminati..

METODE

Kerangka penelitian ini merupakan langkah-langkah yang dilakukan dalam rangka penyelesaian masalah yang akan dibahas.



Gambar 1. Kerangka Penelitian

a. Mendefinisikan Masalah

Mendefinisikan masalah dimana pada objek penelitian ini ingin menerapkan metode clustering untuk menentukan jenis asuransi yang paling diminati.

b. Menganalisa Masalah

Menganalisa masalah merupakan suatu langkah yang dilakukan untuk alur selanjutnya dari penelitian, sehingga dalam penelitian dalam memahami permasalahan yang ada.

c. Mempelajari Literatur

Mempelajari literature berupa penelitian jurnal terdahulu tentang Data Mining dan Clustering.

d. Mengumpulkan Data

Dilakukan terhadap cabang PT. AIA Financial kota padang beserta staff IT PT. AIA Financial kota padang untuk mengetahui secara dalam mengenai jenis asuransi yang paling diminati. Survey penelitian dilakukan dengan minta izin untuk pengambilan data mengenai data produk pada PT. AIA Financial selama satu tahun sebelumnya.

e. Melakukan Analisa Data

Untuk menganalisis data tersebut, maka akan dilakukan suatu analisa mengenai suatu permasalahan yang akan dibahas dalam penelitian ini. Sesuai dengan batasan masalah yang telah ditentukan. Analisa pada penelitian ini dimulai dari menganalisa data yang digunakan seperti nama produk, jenis asuransi, banyak nasabah yang membeli produk.

f. Mengimplementasikan Algoritma K-Means

Pada tahapan ini akan dilakukan tentang penerapan metode algoritma K-means untuk mengetahui jenis asuransi paling diminati oleh konsumen.

g. Melakukan Pengujian

Analisa yang telah dilakukan dengan menggunakan software yang telah ditentukan, untuk mengetahui apakah sistem tersebut memiliki hasil yang sama dengan analisa secara manual, sehingga hasil yang didapatkan sesuai dengan yang diharapkan dengan melakukan perbandingan sistem dengan pencarian manual menghasilkan hasil yang sama.

HASIL DAN PEMBAHASAN

Setelah dilakukan proses data selection diatas, yaitu dengan mengurangi atribut yang tidak diperlukan untuk memudahkan pada saat melakukan tahapan data mining selanjutnya. Dari proses tersebut maka didapatkan hasilnya menjadi atribut nama produk dan atribut jumlah terjual yang dapat dilihat pada gambar berikut:

No	Nama Product	Jumlah Terjual
1	AIA All In One	0
2	Aktivita Prima	0
3	Hassanah Berkah	0
4	AIA Family First Protection	139
5	AIA Infinite Link Assurance	0
6	AIA Golden Harvest Assurance	0
7	AIA Life Secure	18
8	AIA Prosperity Assurance	0
9	AIA Sakinah Assurance	0
10	AIA Infinite Plus Assurance	0
11	AIA Priority Plus Assurance	0
12	AIA Wealth Assurance	0
13	AIA Prolink Assurance	67
14	APL - Plus Basic	7
15	Solusi Prima	1

Gambar 2. Data Setelah Proses Data Selection

1. Menentukan Jumlah Cluster

Jumlah Cluster pada pengolahan data produk asuransi diberikan tiga Cluster yaitu Tidak Diminati, Kurang Diminati, dan Banyak Diminati.

2. Menentukan Titik Pusat Cluster/Centroid

Pusat awal Cluster (Centroid) ditentukan secara random atau acak dari data yang ada. Adapun centroid pada data jenis produk asuransi yang diambil adalah sebagai berikut :

Pusat C1 (Centroid 1) adalah 1

Pusat C2 (Centroid 2) adalah 65

Pusat C3 (Centroid 3) adalah 137

3. Menghitung Jarak Centroid

Formula yang dipakai untuk menghitung jarak antar titik Centroid dengan titik tiap objek yaitu menggunakan Euclidian Distance. Adapun perhitungan Centroid secara manual dengan rumus :

1. Menghitung jarak dari data ke-1 terhadap pusat Cluster

$$D_{11} = \sqrt{((0-1))^2} = 1$$

$$D_{21} = \sqrt{((0-65))^2} = 65$$

$$D_{31} = \sqrt{((0-137))^2} = 137$$

2. Menghitung jarak dari data ke-2 terhadap pusat Cluster

$$D_{12} = \sqrt{((0-1))^2} = 1$$

$$D_{22} = \sqrt{((0-65))^2} = 65$$

$$D_{32} = \sqrt{((0-137))^2} = 137$$

3. Menghitung jarak dari data ke-3 terhadap pusat Cluster

$$D_{13} = \sqrt{((0-1))^2} = 1$$

$$D_{23} = \sqrt{((0-65))^2} = 65$$

$$D_{33} = \sqrt{((0-137))^2} = 137$$

4. Menghitung jarak dari data ke-4 terhadap pusat Cluster

$$D_{14} = \sqrt{((139-1))^2} = 138$$

$$D_{24} = \sqrt{((139-65))^2} = 74$$

$$D_{34} = \sqrt{((139-137))^2} = 2$$

5. Menghitung jarak dari data ke-5 terhadap pusat Cluster

$$D_{15} = \sqrt{((0-1))^2} = 1$$

$$D_{25} = \sqrt{((0-65))^2} = 65$$

$$D_{35} = \sqrt{((0-137))^2} = 137$$

6. Menghitung jarak dari data ke-6 terhadap pusat Cluster

$$D_{16} = \sqrt{((0-1))^2} = 1$$

$$D_{26} = \sqrt{((0-65))^2} = 65$$

$$D_{36} = \sqrt{((0-137))^2} = 137$$

7. Menghitung jarak dari data ke-7 terhadap pusat Cluster

$$D_{17} = \sqrt{((18-1))^2} = 17$$

$$D_{27} = \sqrt{((18-65))^2} = 47$$

$$D_{37} = \sqrt{((18-137))^2} = 119$$

8. Menghitung jarak dari data ke-8 terhadap pusat Cluster

$$D_{18} = \sqrt{((0-1))^2} = 1$$

$$D_{28} = \sqrt{((0-65))^2} = 65$$

$$D_{38} = \sqrt{((0-137))^2} = 137$$

9. Menghitung jarak dari data ke-9 terhadap pusat Cluster

$$D_{19} = \sqrt{((0-1))^2} = 1$$

$$D_{29} = \sqrt{((0-65))^2} = 65$$

$$D_{39} = \sqrt{((0-137))^2} = 137$$

10. Menghitung jarak dari data ke-10 terhadap pusat Cluster

$$D_{110} = \sqrt{((0-1))^2} = 1$$

$$D_{210} = \sqrt{((0-65))^2} = 65$$

$$D_{310} = \sqrt{((0-137))^2} = 137$$

11. Menghitung jarak dari data ke-11 terhadap pusat Cluster

$$D_{111} = \sqrt{((0-1))^2} = 1$$

$$D_{211} = \sqrt{((0-65))^2} = 65$$

$$D_{311} = \sqrt{((0-137))^2} = 137$$

12. Menghitung jarak dari data ke-12 terhadap pusat Cluster

$$D_{112} = \sqrt{((0-1))^2} = 1$$

$$D_{212} = \sqrt{((0-65))^2} = 65$$

$$D_{312} = \sqrt{((0-137))^2} = 137$$

13. Menghitung jarak dari data ke-13 terhadap pusat Cluster

$$D_{113} = \sqrt{((67-1))^2} = 66$$

$$D_{213} = \sqrt{((67-65))^2} = 2$$

$$D_{313} = \sqrt{((67-137))^2} = 70$$

14. Menghitung jarak dari data ke-14 terhadap pusat Cluster

$$D_{114} = \sqrt{((7-1))^2} = 6$$

$$D_{214} = \sqrt{((7-65))^2} = 58$$

$$D_{314} = \sqrt{((7-137))^2} = 130$$

15. Menghitung jarak dari data ke-15 terhadap pusat Cluster

$$D_{115} = \sqrt{((1-1))^2} = 0$$

$$D_{215} = \sqrt{((1-65))^2} = 64$$

$$D_{315} = \sqrt{((1-137))^2} = 136$$

Mengelompokan data berdasarkan jarak terdekat antara data dengan Centroid. Perhitungan dilakukan sebanyak data sampel dan nilai terkecil diberi kode "1". Adapun hasil perhitungan yang dilakukan sebagai berikut :

Tabel 4.2 Perhitungan Jarak dan Pengelompokan Data Iterasi 1

No	Nama Product	Jumlah Terjual (x)	c1	c2	c3	C1	C2	C3
1	AIA All In One	0	1	65	137	1		
2	Aktiviah Prima	0	1	65	137	1		
3	Hassanah Berkah	0	1	65	137	1		

4	AIA Family First Protection	139	138	74	2		1
5	AIA Infinite Link Assurance	0	1	65	137	1	
6	AIA Golden Harvest Assurance	0	1	65	137	1	
7	AIA Life Secure	18	17	47	119	1	
8	AIA Prosperity Assurance	0	1	65	137	1	
9	AIA Sakinah Assurance	0	1	65	137	1	
10	AIA Infinite Plus Assurance	0	1	65	137	1	
11	AIA Priority Plus Assurance	0	1	65	137	1	
12	AIA Wealth Assurance	0	1	65	137	1	
13	AIA Prolink Assurance	67	66	2	70		1
14	APL - Plus Basic	7	6	58	130	1	
15	Solusi Prima	1	0	64	136	1	

Dari tabel iterasi 1 diatas dapat diketahui bahwa :

C1 = Data hasil perhitungan Cluster 1 memiliki 13 anggota yaitu data 1, data 2, data 3, data 5, data 6, data 7, data 8, data 9, data 10, data 11, data 12, data 14, dan data 15.

C2 = Data hasil perhitungan Cluster 2 memiliki 1 anggota yaitu data 13.

C3 = Data hasil perhitungan Cluster 3 memiliki 1 anggota yaitu data 4.

5. Tahap selanjutnya melakukan iterasi 2, dengan menentukan titik Centroid baru dengan menghitung nilai rata-rata dari data yang ada pada Centroid yang sama. Adapun perhitungan Centroid baru dengan menggunakan rumus sebagai berikut :

Centroid 1:

$$c1 = 26/13 = 2$$

Centroid 2:

$$c2 = 67/1 = 67$$

Centroid 3:

$$c3 = 139/1 = 139$$

6. Menghitung jarak Centroid dengan titik Centroid baru. Formula yang dipakai untuk menghitung jarak antar titik Centroid dengan titik tiap objek yaitu menggunakan Euclidian Distance. Adapun pergitungan Centroid secara manual dengan rumus :

1. Menghitung jarak dari data ke-1 terhadap pusat Cluster

$$D_{11} = \sqrt{((0-2))^2} = 2$$

$$D_{21} = \sqrt{((0-67))^2} = 67$$

$$D_{31} = \sqrt{((0-139))^2} = 139$$

2. Menghitung jarak dari data ke-2 terhadap pusat Cluster

$$D_{12} = \sqrt{((0-2))^2} = 2$$

$$D_{22} = \sqrt{((0-67))^2} = 67$$

$$D_{32} = \sqrt{((0-139))^2} = 139$$

3. Menghitung jarak dari data ke-3 terhadap pusat Cluster

$$D_{13} = \sqrt{((0-2))^2} = 2$$

$$D_{23} = \sqrt{((0-67))^2} = 67$$

$$D_{33} = \sqrt{((0-139))^2} = 139$$

4. Menghitung jarak dari data ke-4 terhadap pusat Cluster

$$D_{14} = \sqrt{((139-2))^2} = 137$$

$$D_{24} = \sqrt{((139-67))^2} = 72$$

$$D_{34} = \sqrt{((139-139))^2} = 0$$

5. Menghitung jarak dari data ke-5 terhadap pusat Cluster

$$D_{15} = \sqrt{((0-2))^2} = 2$$

$$D_{25} = \sqrt{((0-67))^2} = 67$$

$$D_{35} = \sqrt{((0-139))^2} = 139$$

6. Menghitung jarak dari data ke-6 terhadap pusat Cluster

$$D_{16} = \sqrt{((0-2))^2} = 2$$

$$D_{26} = \sqrt{((0-67))^2} = 67$$

$$D_{36} = \sqrt{((0-139))^2} = 139$$

7. Menghitung jarak dari data ke-7 terhadap pusat Cluster

$$D_{17} = \sqrt{((18-2))^2} = 16$$

$$D_{27} = \sqrt{((18-67))^2} = 49$$

$$D_{37} = \sqrt{((18-139))^2} = 121$$

8. Menghitung jarak dari data ke-8 terhadap pusat Cluster

$$D_{18} = \sqrt{((0-2))^2} = 2$$

$$D_{28} = \sqrt{((0-67))^2} = 67$$

$$D_{38} = \sqrt{((0-139))^2} = 139$$

9. Menghitung jarak dari data ke-9 terhadap pusat Cluster

$$D_{19} = \sqrt{((0-2))^2} = 2$$

$$D_{29} = \sqrt{((0-67))^2} = 67$$

$$D_{39} = \sqrt{((0-139))^2} = 139$$

10. Menghitung jarak dari data ke-10 terhadap pusat Cluster

$$D_{110} = \sqrt{((0-2))^2} = 2$$

$$D_{210} = \sqrt{((0-67))^2} = 67$$

$$D_{310} = \sqrt{((0-139))^2} = 139$$

11. Menghitung jarak dari data ke-11 terhadap pusat Cluster

$$D_{111} = \sqrt{((0-2))^2} = 2$$

$$D_{211} = \sqrt{((0-67))^2} = 67$$

$$D_{311} = \sqrt{((0-139))^2} = 139$$

12. Menghitung jarak dari data ke-12 terhadap pusat Cluster

$$D_{112} = \sqrt{((0-2))^2} = 2$$

$$D_{212} = \sqrt{((0-67))^2} = 67$$

$$D_{312} = \sqrt{((0-139))^2} = 139$$

13. Menghitung jarak dari data ke-13 terhadap pusat Cluster

$$D_{113} = \sqrt{((67-2))^2} = 65$$

$$D_{213} = \sqrt{((67-67))^2} = 0$$

$$D_{313} = \sqrt{((67-139))^2} = 72$$

14. Menghitung jarak dari data ke-14 terhadap pusat Cluster

$$D_{114} = \sqrt{((7-2))^2} = 5$$

$$D_{214} = \sqrt{((7-67))^2} = 60$$

$$D_{314} = \sqrt{((7-139))^2} = 132$$

15. Menghitung jarak dari data ke-15 terhadap pusat Cluster

$$D_{115} = \sqrt{((1-2))^2} = 1$$

$$D_{215} = \sqrt{((1-67))^2} = 66$$

$$D_{315} = \sqrt{((1-139))^2} = 138$$

7. Mengelompokan data berdasarkan jarak terdekat antara data dengan Centroid. Perhitungan dilakukan sebanyak data sampel dan nilai terkecil diberi kode "1". Adapun hasil perhitungan yang dilakukan sebagai berikut :

Tabel 2. Perhitungan Jarak dan Pengelompokan Data Iterasi 2

No	Nama Product	Jumlah Terjual (x)	c1	c2	c3	C1	C2	C3
1	AIA All In One	0	2	67	139	1		
2	Aktivita Prima	0	2	67	139	1		
3	Hassanah Berkah	0	2	67	139	1		
4	AIA Family First Protection	139	137	72	0			1
5	AIA Infinite Link Assurance	0	2	67	139	1		
6	AIA Golden Harvest Assurance	0	2	67	139	1		
7	AIA Life Secure	18	16	49	121	1		
8	AIA Prosperity Assurance	0	2	67	139	1		
9	AIA Sakinah Assurance	0	2	67	139	1		
10	AIA Infinite Plus Assurance	0	2	67	139	1		
11	AIA Priority Plus Assurance	0	2	67	139	1		
12	AIA Wealth Assurance	0	2	67	139	1		
13	AIA Prolink Assurance	67	65	0	72		1	
14	APL - Plus Basic	7	5	60	132	1		
15	Solusi Prima	1	1	66	138	1		

Dari tabel iterasi 2 diatas dapat diketahui bahwa :

C1 = Data hasil perhitungan Cluster 1 memiliki 13 anggota yaitu data 1, data 2, data 3, data 5, data 6, data 7, data 8, data 9, data 10, data 11, data 12, dan data 15.

C2 = Data hasil perhitungan Cluster 2 memiliki 1 anggota yaitu data 13.

C3 = Data hasil perhitungan Cluster 3 memiliki 1 anggota yaitu data 4.

Dikarenakan hasil iterasi 1 dan iterasi 2 memiliki kesamaan posisi Cluster, maka proses dihentikan. Kesimpulan dari hasil perhitungan yang dilakukan mulai iterasi 1 sampai iterasi 2 adalah pada literasi 1, cluster 1(c1) memiliki 13 anggota yaitu data 1, data 2, data 3, data 5, data 6, data 7, data 8, data 9, data 10, data 11, data 12, data 14, dan data 15, dimana c1 merupakan kategori jenis produk asuransi yang tidak diminati. Pada Cluster 2(c2) memiliki 1 anggota yaitu data ke-13, dimana c2 merupakan kategori jenis produk asuransi yang kurang diminati. Sedangkan pada cluster 3(c3) memiliki 1 anggota yaitu data ke-4, dimana c3 merupakan kategori jenis produk asuransi yang banyak diminati.

KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan oleh penulis pada PT. AIA Financial dalam mengelompokkan jenis produk asuransi yang paling banyak diminati dan yang kurang diminati dan pengujian data dilakukan dengan software RapidMiner 9.2 , dapat diambil kesimpulannya sebagai berikut :

1. Teknik Data Mining dengan Algoritma K-Means Clustering dapat diterapkan di PT. AIA Financial untuk mengetahui produk asuransi mana yang paling banyak diminati oleh nasabah.
2. Hasil dari pengujian data 2 tahun terakhir yaitu 2017-2018, diperoleh 3 pola yaitu tidak diminati, kurang diminati, dan banyak diminati. Dimana pada katerogi produk asuransi yang tidak diminati ada 13 jenis produk yaitu produk asuransi AIA All In One, Aktivia Prima, Hassanah Berkah, AIA Infinite Link Assurance, AIA Golden Harvest Assurance, AIA Life Secure, AIA Prosperity Assurance, AIA Sakinah Assurance, AIA Infinite Plus Assurance, AIA Wealth Assurance, AIA Priority Plus Assurance, APL-Plus Basic, dan Solusi Prima. Sedangkan pada kategori kurang diminati terdapat 1 jenis produk yaitu AIA Prolink. Dan untuk kategori banyak diminati terdapat 1 jenis produk yaitu AIA Family First Protection.
3. Dari pengujian yang dilakukan pada data penjualan produk asuransi banyak diperoleh pengetahuan baru yang dapat digunakan oleh PT. AIA Financial untuk meningkatkan penjualan produknya.

DAFTAR PUSTAKA

- Sutrisno, S., Afriyudi, A., & Widiyanto, W. (2013). ***Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering Study Kasus Pt. Indomarco Palembang***. JURNAL MAHASISWA TI S1.
- Normah, N., Nurajizah, S., & Salbinda, A. (2021). ***Penerapan data mining metode k-means clustering untuk analisa penjualan pada toko fashion hijab Banten***. Jurnal Khatulistiwa Informatika, 7(2), 158-163.
- Metisen, B. M., & Sari, H. L. (2015). ***Analisis clustering menggunakan metode K-Means dalam pengelompokkan penjualan produk pada Swalayan Fadhila***. Jurnal media infotama, 11(2).
- Nabila, Z., Isnain, A. R., Permata, P., & Abidin, Z. (2021). ***Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means***. Jurnal Teknologi Dan Sistem Informasi, 2(2), 100-108.
- Sadewo, M. G., Windarto, A. P., & Hartama, D. (2017). ***Penerapan datamining pada populasi daging ayam ras pedaging di indonesia berdasarkan provinsi menggunakan k-means clustering***. InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan), 2(1), 60-67.
- Sutrisno, S., Afriyudi, A., & Widiyanto, W. (2013). ***Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering Study Kasus Pt. Indomarco Palembang***. JURNAL MAHASISWA TI S1.

- Handoko, S., Fauziah, F., & Handayani, E. T. E. (2020). ***Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering***. Jurnal Ilmiah Teknologi dan Rekayasa, 25(1), 76-88.
- Hendrian, S. (2018). ***Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan***. Faktor Exacta, 11(3).
- Tamba, S. P., & Kesuma, F. T. (2019). ***Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering: data mining; k-means-clustering***. Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA), 2(2), 67-72.