

## Terakreditasi SINTA Peringkat 4

Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 28/E/KPT/2019  
masa berlaku mulai Vol.3 No. 1 tahun 2018 s.d Vol. 7 No. 1 tahun 2022

Terbit online pada laman web jurnal:  
<http://publishing-widyagama.ac.id/ejournal-v2/index.php/jointecs>



Vol. 6 No. 3 (2021) 153 - 160

# JOINTECS

## (Journal of Information Technology and Computer Science)

e-ISSN:2541-6448

p-ISSN:2541-3619

### Analisa *Cluster* Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 dengan Algoritma *K-means*

Iis Setiawan Mangku Negara<sup>1</sup>, Purwono<sup>2</sup>, Imam Ahmad Ashari<sup>3</sup>

<sup>1,3</sup>Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Harapan Bangsa

<sup>2</sup>Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Harapan Bangsa

<sup>1</sup>iissmn@uhb.ac.id, <sup>2</sup>purwono@uhb.ac.id, <sup>3</sup>imamahmadashari@uhb.ac.id

#### Abstract

*Covid-19 has had a negative impact on the economic sector in Indonesia. This can be seen from the losses experienced by industry players in the form of a decrease in income turnover. Sales strategy needs to be done so that losses can be minimized. Sales transaction analysis can be done to find product groups with the most sales data so that stock management can be fulfilled and increase sales transactions. Berkah Abadi Minimarket is an industry that has been affected by this pandemic. Data analysis has not been carried out to find out which product has the most sales data, so it is necessary to analyze it with the k-means algorithm. This algorithm can group data based on similar characteristics. The application of the algorithm on 278480 transaction data, obtained three sales data clusters, namely cluster 2 or the highest sales of 57 products, cluster 1 or moderate sales of 57 products and the rest are cluster 0 with low sales. The result of the accuracy of the clustering model generated by the confusion matrix is 87%. Based on these results, the owners of the Berkah Abadi Minimarket can be helped in making decisions on stock management while the Covid-19 pandemic is still ongoing.*

Keywords: *analysis; clustering; transactions; covid-19; data mining.*

#### Abstrak

Covid-19 berdampak buruk pada sektor ekonomi di Indonesia. Hal ini terlihat dari kerugian yang dialami pelaku industri berupa penurunan omset penghasilan. Strategi penjualan perlu dilakukan agar kerugian dapat diminimalisir. Analisa transaksi penjualan bisa dilakukan untuk menemukan kelompok produk dengan data penjualan terbanyak sehingga manajemen stok dapat terpenuhi dan meningkatkan transaksi penjualan. Minimarket Berkah Abadi merupakan industri yang terdampak oleh pandemi ini. Analisa data belum dilakukan untuk mencari tahu produk mana yang memiliki data penjualan terbanyak, sehingga perlu dilakukan analisa dengan algoritma *k-means*. Algoritma ini dapat mengelompokkan data berdasarkan kemiripan karakteristiknya. Penerapan algoritma pada 278480 data transaksi, didapatkan tiga *cluster* data penjualan yaitu *cluster* 2 atau penjualan terbanyak sebanyak 57 produk, *cluster* 1 atau penjualan sedang sebanyak 5 produk dan sisanya adalah *cluster* 0 dengan penjualan rendah. Hasil akurasi model klusterisasi yang dihasilkan dengan *confusion matrix* adalah 87%. Berdasarkan hasil ini pemilik Minimarket Berkah Abadi dapat terbantu dalam membuat keputusan pada manajemen stok barang pada saat pandemi Covid-19 masih berlangsung.

**Kata kunci:** analisa; klusterisasi; transaksi; covid-19; data mining.

© 2021 Jurnal JOINTECS

#### 1. Pendahuluan

Covid-19 telah menjadi pandemi global dan masih terus mengalami kenaikan angka infeksi harian di Indonesia

[1]. Akibat pandemi ini, banyak negara-negara di dunia termasuk Indonesia terkena dampaknya. Sektor ekonomi menjadi salah satu bidang yang cukup serius terdampak

Diterima Redaksi : 23-08-2021 | Selesai Revisi : 19-09-2021 | Diterbitkan Online : 30-09-2021

buruk [2]. Sebagai contoh paling nyata adalah adanya penurunan omzet penghasilan pelaku UMKM. Selain itu juga pasar ritel modern, skala besar, mikro hingga kecil mulai mengalami penurunan penghasilan [3]. Dampak ini menuntut adanya strategi khusus untuk meningkatkan penjualan selama pandemi. Peningkatan penjualan dapat diketahui dengan dilakukannya analisa data berupa produk mana yang memiliki transaksi penjualan terbanyak selama pandemi.

Pelaku usaha harus melakukan analisa data transaksi untuk mencari tahu produk mana yang paling banyak dibutuhkan masyarakat selama pandemi. Aplikasi kasir (*point of sales*) telah diterapkan pada usaha pertokoan baik yang skala kecil maupun besar. Aplikasi kasir menghasilkan data transaksi penjualan selama periode tertentu seperti data harian, mingguan, bulanan hingga tahunan. Aplikasi kasir menyimpan data transaksi berupa jumlah transaksi dan total penjualan dari setiap produk. Data disimpan dalam *database* aplikasi mereka namun dalam kurun waktu tertentu data dibiarkan menumpuk bahkan dihapus karena dianggap sebagai sampah atau memenuhi media penyimpanan saja [4]. Hal ini terjadi dikarenakan belum pemahannya pemilik usaha bahwa mereka dapat menemukan pola tertentu untuk mempermudah dalam strategi penjualan pada periode selanjutnya.

Mini Market Berkah Abadi Tegal merupakan salah satu usaha yang bergerak dibidang penjualan barang atau ritel. Saat ini data transaksi penjualan yang dihasilkan belum dimanfaatkan dengan baik untuk pengambilan keputusan penjualan. Usaha ini masih mengelola data secara manual dan belum dilakukan analisa secara menyeluruh yang berakibat pada kesalahan dalam penentuan stok barang sehingga banyak barang yang menumpuk karena kurang laku serta banyak stok barang laku yang kosong. Hal ini tentunya merugikan Mini Market Berkah Abadi Tegal karena berdampak pada ruginya usaha dan belum mampunya memenuhi kebutuhan masyarakat selama pandemi karena kesalahan dalam pengambilan keputusan.

Kumpulan data transaksi dalam jumlah besar sebetulnya dapat dimanfaatkan untuk keperluan prediksi dalam menentukan strategi penjualan selanjutnya dengan metode *Data Mining* [5]. Metode ini dapat digunakan untuk menghasilkan pola-pola yang valid, berpotensi memiliki manfaat serta dapat dipahami dengan mudah dari kumpulan data dalam jumlah besar [6]. Dalam bidang penjualan *data mining* bisa digunakan untuk mengetahui kelompok produk yang dibeli oleh konsumen berdasarkan pola transaksinya sehingga dapat dijadikan sebagai cara untuk menentukan rencana stok produk hingga teknik promosi [7].

Salah satu metode *data mining* yang dapat diimplementasikan dalam data penjualan yaitu teknik *clustering*. Teknik ini dapat kita manfaatkan untuk mengelompokkan data-data tertentu berdasarkan kemiripan polanya [8]. Algoritma *K-means* bisa

digunakan sebagai metode yang mampu melakukan klusterisasi data berdasarkan pola-pola distribusi data, sehingga ditemukannya keterkaitan dari setiap atributnya [9].

Penelitian terkait *data mining* dalam klusterisasi dengan algoritma *k-means* pernah dilakukan oleh Rozi [10], dimana hasilnya adalah terdapat *cluster* penjualan sepeda motor yang dijual secara murah, normal dan mahal. Penelitian [10] menghasilkan nilai akurasi model klusterisasi sebesar 81%. Penelitian yang dilakukan oleh Handoko [11] dimana *data mining* dengan metode *k-means* menghasilkan akurasi sebesar 99% pada model klusterisasi yang digunakan untuk melihat pola daerah penjualan produk yang tinggi, sedang dan rendah pada penjualan paket data Telkomsel. Penelitian yang dilakukan oleh Devin [12] memanfaatkan algoritma *naïve bayes* dan *k-means* untuk *sentiment analysis* pada opini penjualan paket internet. Penelitian [12] menghasilkan akurasi sebesar 88,3% pada model klusterisasi *k-means*. Penelitian terkait juga pernah dilakukan oleh Ahmad [13] dimana teknik *data mining* dengan *clustering* dapat menghasilkan pola kinerja karyawan tingkat kinerja sangat baik, tingkat kinerja baik, tingkat kinerja cukup dan tingkat kinerja kurang baik pada perusahaan nasional. Penelitian [13] menghasilkan nilai akurasi model klusterisasi sebesar 56%.

Kontribusi kami dalam penelitian ini ialah melakukan analisa untuk menemukan sebuah pola tertentu dari data transaksi penjualan yang dihasilkan oleh aplikasi kasir milik Mini Market Berkah Abadi Tegal selama Pandemi Covid-19. Hasil yang diharapkan adalah adanya kelompok data yang berpengaruh pada pengambilan keputusan manajemen stok barang yang akan dijual. Keputusan ini akan meminimalisir kerugian yang sebelumnya diterima oleh Mini Market Berkah Abadi Tegal serta dapat menjawab kebutuhan masyarakat selama pandemi covid-19 berlangsung.

## 2. Metode Penelitian

Metode penelitian merupakan langkah-langkah yang dilakukan oleh peneliti untuk mengumpulkan data penelitian. Data ini diolah dan dianalisis secara ilmiah hingga mencapai tujuan penelitian. Langkah-langkah dalam penelitian ini antara lain (1) Pengumpulan Data, (2) Pemilihan Data, (3) Transformasi Data, (4) Metode Elbow dan (5) Penerapan Algoritma *K-means* untuk klusterisasi.

### 2.1. Pengumpulan Data

Penelitian ini dilakukan di Mini Market Berkah Abadi Tegal. Kami melakukan observasi dan wawancara secara langsung pada pemilik dan operator aplikasi kasir. Data yang berhasil dikumpulkan adalah data transaksi penjualan dalam periode satu tahun dari bulan Mei 2020 hingga Mei 2021 atau selama 1 tahun pandemi covid-19. Kami mendapatkan data transaksi penjualan

sebanyak 278480 *record* yang tersimpan pada *database* aplikasi kasir. Pada awalnya data berformat SQL (*Structured Query Language*) karena aplikasi kasir menggunakan penyimpanan data dengan *database* MySQL. Operator selanjutnya melakukan ekspor data dalam dengan format *csv* menggunakan bantuan *software* Xampp yaitu *phpmyadmin*. Data dalam format *csv* dapat memudahkan dalam proses *data mining*.

## 2.2. Pemilihan Data

Data yang dikumpulkan belum sepenuhnya menjadi data yang siap digunakan dalam proses klusterisasi. Tahap pemilihan data digunakan untuk mengurangi kompleksitas atribut yang akan dikelola pada tahap transformasi data dan analisis [14]. Pada tahap pemilihan data, data yang dipakai adalah transaksi penjualan yang terjadi pada periode satu tahun dengan pemilihan *field-field* seperti kode transaksi, nama barang, qty, harga awal dan laba pada Tabel 1.

Tabel 1. Pemilihan Data Transaksi Penjualan

Kode Transaksi	Nama Barang	Qty	Harga Awal	Laba
202007160001	Kacang Kiloan	1	7000	750
202007190013	Cornetto Mini All Varian 28ml	3	7500	1251
202007190013	Yupi Kiss Strawberry	2	1000	280
202007190013	Chocolatos Choc Drink Isi 5	1	8500	750
202007190013	Keripik Pisang Madu Kiloan	1	7000	800
....	....	....	....	....
202106170050	Beras Ketan 1 Kg	1	12000	1000

## 2.3. Transformasi Data

Tahap transformasi data atau bisa disebut dengan *pra processing* adalah proses perubahan data menjadi data yang mampu diolah dengan algoritma *clustering*. Bentuk data dapat bersifat numerik, klasifikasi dan sebagainya [15]. Transformasi menghasilkan fitur-fitur yang akan diseleksi (*feature selection*) agar proses algoritma *k-means* dapat berjalan lebih optimal.

## 2.4. Seleksi Fitur

Seleksi fitur (*feature selection*) merupakan tahapan yang digunakan untuk mereduksi dimensi dari fitur yang digunakan. Keuntungan dari seleksi fitur adalah algoritma dapat bekerja dengan performa yang lebih baik, membentuk model yang mudah dipahami, dapat menghemat waktu dan memori saat pemrosesan, serta dapat memilih subset fitur terbaik yang dapat memberikan akurasi tertinggi [16]. Kami menggunakan pendekatan *Min-Max Normalization* untuk mengoptimasi fitur yang akan digunakan. *Min-Max Normalization* merupakan metode normalisasi dengan

melakukan transformasi linier terhadap data asli sehingga tercipta keseimbangan nilai perbandingan antara data sebelum dan sesudah proses [17].

## 2.5. Metode Elbow

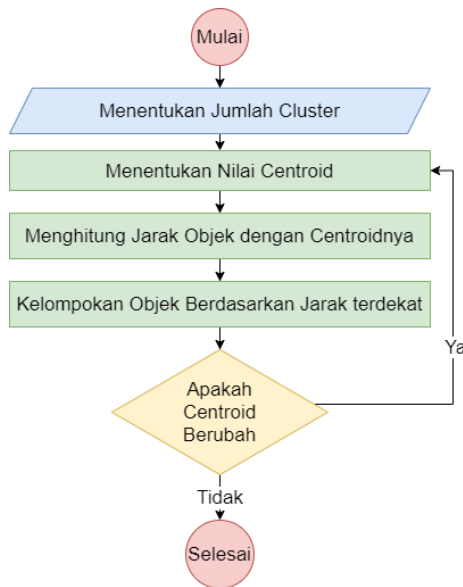
Sebelum melakukan implementasi pada algoritma *k-means* kita dapat terlebih dahulu mencari tahu jumlah *k cluster* terbaik dengan metode *elbow*. Metode ini bertujuan untuk memilih nilai *k* yang kecil serta memiliki nilai *withinss* yang rendah [18]. Penentuan jumlah *cluster* optimal dilakukan dengan memperhatikan nilai SSE (*Sum Squared Error*) yaitu perbandingan dari perhitungan jumlah kuadrat jarak antara *centroid* dengan anggota *cluster* antara jumlah *cluster* yang akan membentuk siku pada suatu titik [19]. Perhitungan nilai SSE dapat dilakukan dengan rumus 1.

$$SSE = \sum_{K=1}^K \sum_{x_i \in C_k} ||x_i - \mu_k||^2 \quad (1)$$

Berdasarkan rumus 1, *k* merupakan jumlah *cluster* yang digunakan, pada rumus 1 *k cluster* dimulai dengan nilai 1. Penjumlahan  $x_i$  merupakan nilai data ke - i pada keseluruhan *cluster* misalkan *cluster* 1, *cluster* 2 dan seterusnya. Nilai  $\mu_k$  ialah *centroid* dari *cluster* atau nilai mean dari total *dataset cluster*. Semakin besar jumlah *cluster k* maka nilai SSE akan semakin kecil [20]. Langkah-langkah yang digunakan pada metode *elbow* meliputi (1) Inisialisasi awal nilai *K*, (2) Menaikan nilai *K*, (3) Menghitung hasil SSE dari tiap nilai *K*, (4) Melihat hasil SSE dari nilai *K* yang turun secara drastis dan (5) Menetapkan nilai *K* yang berbentuk siku.

## 2.6. K-means Clustering

*Clustering* atau analisis *cluster* merupakan sebuah proses pembentukan kelompok data berbentuk *cluster* dari himpunan data yang sebelumnya belum diketahui kelompok-kelompoknya [21]. Data transaksi yang telah dilakukan transformasi selanjutnya akan di klusterisasi dengan salah satu algoritma *data mining* yaitu *k-means clustering*. Algoritma ini menggunakan pemodelan tanpa supervisi dan pengelompokan datanya dilakukan secara partisi. Data dikelompokkan dalam beberapa kelompok dan setiap kelompok memiliki karakteristik yang mirip satu sama lain namun dengan kelompok lainya memiliki karakteristik yang berbeda yang bertujuan untuk memilimalisir perbedaan setiap data dalam sebuah *cluster* dan memaksimalkan perbedaan dengan *cluster* lain [22]. Dalam algoritma *k-means clustering* terdapat beberapa istilah-istilah seperti *cluster*; *centroid* dan iterasi. Menurut Muliono [23] *cluster* yaitu kelompok atau grup data, *centroid* yaitu titik pusat dari sebuah *cluster* dan iterasi yaitu pengulangan proses, akan berhenti ketika hasil iterasi telah konvergen. Berikut ini merupakan tahapan-tahapan yang digambarkan dalam *flowchart* dari algoritma *k-means clustering* yang dapat dilihat secara jelas pada Gambar 1.



Gambar 1. Flowchart Algoritma K-Means.

Berdasarkan Gambar 1, langkah-langkah yang digunakan dalam algoritma *k-means clustering* [21] dimulai dengan tahap pertama yaitu menentukan jumlah  $k$  sebagai *cluster* yang akan dibentuk. Penentuan banyaknya jumlah *cluster* dilakukan dengan mempertimbangkan faktor teoritis dan konseptual yang kemudian digunakan sebagai *cluster* usulan. Tahap kedua membangkitkan  $k$  *centroid* awal atau titik pusat *cluster* secara *random*. Hal ini dapat dilakukan dalam bentuk acak dari beberapa objek yang tersedia sebanyak jumlah dari  $k$  *cluster* kemudian dihitung *centroid cluster* ke- $i$ . Perhitungan ini dapat dilakukan dengan menggunakan rumus 2.

$$C_k = \frac{1}{n_k} \sum d_i \quad (2)$$

Berdasarkan rumus 2  $C_k$  ialah nilai centroid pada  $k$  *cluster* yang dicari. Nilai *Centroid k* merupakan pembagian antara nilai 1 dengan jumlah data dalam cluster  $k$  ( $n_k$ ). Selanjutnya dilakukan penjumlahan dari nilai jarak yang masuk dalam masing-masing cluster ( $d_i$ ). Tahap ketiga melakukan perhitungan jarak *Euclidean* antara setiap objek ke masing-masing *centroid* dari masing-masing data *cluster*. Jarak antara objek dengan *centroid* juga dihitung pada tahap ini. Perhitungan ini dapat dilihat pada rumus 3 berikut:

$$D(x_i, \mu_i) = \sqrt{\sum_{i=1}^n (x_i, \mu_i)^2} \quad (3)$$

Berdasarkan rumus 3, jarak *Euclidean*  $D(x_i, \mu_i)$  adalah jarak antara cluster  $n$  dengan pusat cluster  $\mu$  pada data ke- $i$  yang merupakan bobot data ke- $i$  pada cluster yang ingin dicari jaraknya,  $\mu_i$  bobot data ke- $i$  pada pusat

*cluster*. Tahap keempat menghitung masing-masing objek ke dalam *centroid* yang paling terdekat. Tahap kelima adalah melakukan iterasi ke- $i$  dan menentukan posisi *centroid* baru dengan menggunakan persamaan sebelumnya. Tahap terakhir melakukan langkah pada poin ketiga jika posisi *centroid* baru tidak sama dengan iterasi atau iterasi akan berhenti jika rasio tidak lebih besar dari nilai rasio sebelumnya hingga hasil perhitungan masing-masing data bersifat konvergen.

## 2.6. Hasil Klusterisasi

Hasil klusterisasi disesuaikan setelah mendapatkan jumlah  $k$  terbaik dari metode *elbow*. Jumlah  $k$  kemudian diimplementasikan pada algoritma *k-means*. Algoritma ini akan menghasilkan kelompok-kelompok produk berbentuk *cluster* dari data transaksi penjualan. Setiap produk akhirnya akan memiliki label klusteranya sendiri. Kluster data ini kemudian dibedakan berdasarkan warna dan titik *centroid*nya. Kluster data ini diharapkan dapat digunakan sebagai penunjang keputusan oleh pembuat kebijakan selama pandemi covid-19 khususnya pada manajemen stok produk yang akan dijual pada minimarket Berkah Abadi Tegal.

## 2.7. Evaluasi Hasil

Evaluasi hasil yang digunakan dalam penelitian ini adalah metode *confusion matrix*. Metode ini menghasilkan nilai *accuracy*, *recall*, *precision* dan *f1-score* dari validasi algoritma terhadap *dataset* yang ada [24]. Menurut Suyanto [25] *accuracy* disebut juga dengan tingkat pengenalan atau rasio prediksi benar dari keseluruhan data. Perhitungan *accuracy* dilakukan dengan rumus 4. *Recall* dapat disebut dengan *true positive rate* atau *sensitivity* dan dapat dihitung dengan rumus 5. *Precision* merupakan ukuran kepastian label data positif yang dapat dihitung dengan rumus 6 dan *f1-score* merupakan rata-rata *harmonic* dari *precision* dan *recall* yang dapat dihitung dengan rumus 7.

$$Accuracy = \frac{TP + TN}{P + N} \quad (4)$$

$$Recall = \frac{TP}{P} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

Berdasarkan rumus 4-7, TP (*True Positive*) yaitu jumlah *tuple* positif dilabeli benar oleh model klusterisasi. TN (*True Negative*) yaitu jumlah *tuple* negatif dilabeli benar oleh model klusterisasi. FP (*False Positive*) yaitu jumlah *tuple* negatif dilabeli salah oleh model klusterisasi. FN (*False Negative*) yaitu jumlah *tuple* positif dilabeli salah oleh model klusterisasi.

### 3. Hasil dan Pembahasan

#### 3.1. Transformasi Data dan Seleksi Fitur

Data yang berhasil dikumpulkan yaitu sebanyak 278480 record data transaksi. Data tersebut dilakukan transformasi agar mudah diproses oleh algoritma clustering yaitu *k-means*. Format hasil dari transformasi data tersebut dapat dilihat pada Tabel 2.

Tabel 2. Hasil Transformasi Data

Nama Barang	Jumlah Transaksi	Total Penjualan
2 Tang Black Tea 1 Renceng (Isi 10)	1	1
2 Tang Melati 1 Renceng (Isi 10)	1	1
Aqua 1500 Ml 1 Dus	10	11
Ribut Kiloan	77	83
Minyak Tawon Ff	8	9
2 Tang Black Tea 1 Renceng (Isi 10)	1	1

Berdasarkan Tabel 2, transformasi data menampilkan label-label khusus yaitu nama barang, jumlah transaksi dan total penjualan. Label jumlah transaksi dan total penjualan kemudian dipilih sebagai fitur yang akan digunakan untuk klusterisasi dengan algoritma *k-means*. Hasil normalisasi fitur dengan metode *min-max normalization* menghasilkan *cluster center* data transaksi berdasarkan jumlah transaksi dan total penjualan. Hasil normalisasi dengan *min-max normalization* dapat dilihat pada Tabel 3.

Tabel 3. Normalisasi Fitur dengan *Min-Max Normalization*

Cluster	Jumlah Transaksi	Total Penjualan
1	0.00452355	0.00406837
2	0.78020446	0.98209805
3	0.15735578	0.2105814

#### 3.2. Metode Elbow

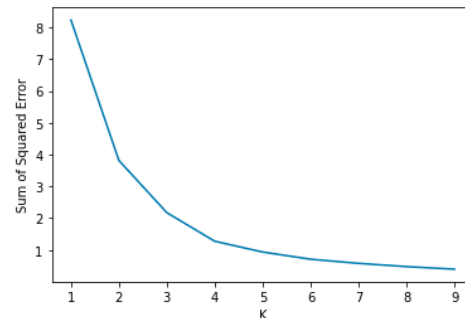
Hasil dari uji coba untuk menentukan jumlah *k cluster* terbaik dilakukan dengan jumlah  $k=2$  hingga  $k=9$ . Parameter yang digunakan adalah melakukan *fit* data antara jumlah transaksi dengan total penjualan. Metode *elbow* yang digunakan pada *range*  $k=2$  hingga  $k=9$  menghasilkan nilai SSE. Nilai SSE yang dihasilkan dengan rumus 1 menunjukkan nilai yang berbeda pada setiap iterasinya. Hasil perhitungan nilai SSE dengan metode *elbow* ditampilkan pada Tabel 4.

Tabel 4. Perhitungan SSE dengan Metode Elbow

Iterasi ke-	Nilai SSE
2	8.233054446326896
3	3.8154096024027044
4	2.173059544143249
5	1.2736086834800011
6	0.9349618202773542
7	0.7089127401197532
8	0.5786496025783243
9	0.4758691998239765

Visualisasi hasil perhitungan SSE dengan metode *elbow* yaitu grafik yang berbentuk siku. Angka-angka ini kemudian divisualisasikan dengan diagram *cartesius*

dengan sumbu  $x$  dan sumbu  $y$ . Sumbu  $x$  mewakili jumlah iterasi *k cluster* yaitu iterasi ke-2 hingga ke-9, sedangkan sumbu  $y$  merupakan nilai SSE yang dihasilkan. Jumlah *k cluster* terbaik dari iterasi ini tepat pada titik keseimbangan sebuah siku. Visualisasi berbentuk siku ini dapat dilihat dengan mudah pada Gambar 2.

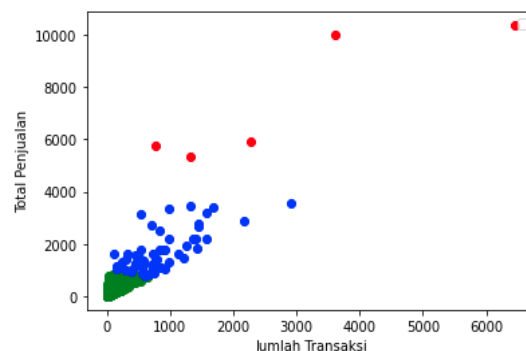


Gambar 2. Penentuan Jumlah *K Cluster* dengan Metode Elbow

Berdasarkan Gambar 2, titik siku pada grafik *elbow* terletak pada  $k=3$ . Titik ini dianggap sebagai jumlah *k cluster* terbaik pada algoritma *k-means clustering*. Nilai  $k=3$  selanjutnya digunakan untuk menghasilkan 3 *cluster* terbaik pada data transaksi Mini Market Berkah Abadi Tegal.

#### 3.3. Hasil Klusterisasi Data Transaksi

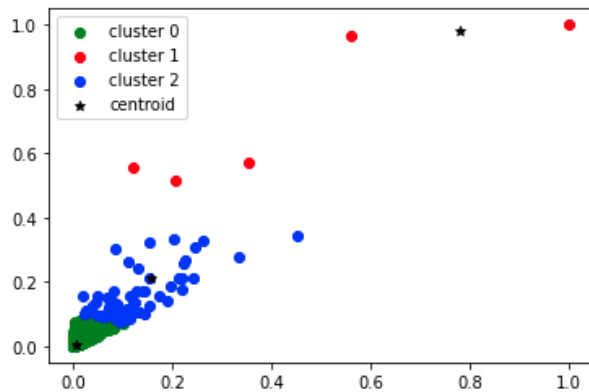
Berdasarkan langkah-langkah yang digunakan dalam melakukan klusterisasi dengan algoritma *k-means clustering*, kami mengaplikasikan nilai  $k = 3$  sebagai jumlah klusternya. Ketiga *cluster* ini akan mengelompokkan tiga kelompok produk serupa berdasarkan pola penjualannya terkait besaran jumlah transaksi terhadap total penjualan. Visualisasi ketiga *cluster* dari data transaksi tersebut dapat dilihat pada Gambar 3. *Cluster* dari setiap kelompok transaksi dibedakan dengan warna hijau, biru dan merah. Sampai pada tahap ini sudah jelas terdapat tiga *cluster*, namun kita belum mengetahui dari setiap warna tersebut masuk pada *cluster* mana. Untuk mengetahuinya, kita akan mendapatkan nilai *centroid* dari setiap *cluster*.



Gambar 3. Tiga Cluster Data Transaksi

Nilai *centroid* dihasilkan dengan menghitung rumus 2. Berdasarkan perhitungan tersebut kami mendapatkan nilai *centroid* atau pusat *cluster* dari setiap kelompok

data transaksi. *Centroid* dilambangkan dengan simbol bintang. Visualisasi dari setiap *centroid* dalam *cluster* data transaksi dapat dilihat pada Gambar 4.



Gambar 4. *Centroid* pada *Cluster* Data Transaksi

Adapun hasil dari klusterisasi data transaksi yaitu *cluster 0*, *cluster 1* dan *cluster 2*. Ketiga *cluster* ini merepresentasikan data-data transaksi penjualan berdasarkan jumlah transaksi yang disandingkan dengan total penjualannya. Pembagian ketiga *cluster* dapat dilihat pada Tabel 5.

Tabel 5. Hasil *Cluster* Data Transaksi

Cluster	Jumlah Transaksi	Total Penjualan
0	0.0	0.0
1	0.2	0.2
2	0.8	1.0

Berdasarkan Tabel 5, *cluster 0* menunjukkan bahwa produk-produk tersebut selama pandemi covid-19 memiliki jumlah transaksi dan total penjualan yang sedikit. *Cluster 1* menunjukkan adanya jumlah transaksi dan total penjualan yang sedang dan *cluster 2* adalah produk-produk dengan jumlah transaksi dan total penjualan yang banyak. Jumlah produk dengan *cluster 2* atau terbanyak berjumlah 57 produk dan untuk *cluster 1* atau sedang berjumlah 5 produk selebihnya adalah *cluster 3* atau sedikit. Produk-produk pada *cluster 0* ditampilkan pada Tabel 6. Produk-produk pada *cluster 1* ditampilkan pada Tabel 7 dan produk-produk pada *cluster 2* ditampilkan pada Tabel 8. Setiap Tabel hanya menampilkan sebanyak 5 produk yang diurutkan berdasarkan total penjualan terbanyak dari setiap *cluster*.

Tabel 6. Produk Pada *Cluster 0*

No	Nama Barang	Jumlah Transaksi	Total Penjualan
1	Golda Coffe 200ml	229	354
2	Mama Lemon Jeruk Nipis 230ml	317	342
3	Kris Bee Kentang Goreng 30g	261	332
4	Teh Javana Gula Batu 350ml	218	301
5	Sedap Ayam Bawang Telur	104	277

Tabel 7. Produk Pada *Cluster 1*

No	Nama Barang	Jumlah Transaksi	Total Penjualan	Cluster
1	Mie Sedap Baso Spesial	229	844	1
2	Milku Coklat 200ml	436	740	1
3	Sedap Goreng Salero Padang	323	654	1
4	Ekonomi Lemon 500k	262	518	1
5	Floridina Orange 350ml	267	428	1

Tabel 8. Produk Pada *Cluster 2*

No	Nama Barang	Jumlah Transaksi	Total Penjualan	Cluster
1	Telor 1/2kg	2919	3556	2
2	2 Tang Hijau 1 Pack	1317	3444	2
3	Gas 3 Kg	1688	3428	2
4	Gula Pasir 1/4 Kg	983	3335	2
5	Obat Nyamuk Kingkong	1585	3214	2

Berdasarkan data dari Tabel 6, 7 dan 8 dapat dilihat dengan jelas terkait adanya perbedaan yang signifikan antara jumlah total penjualan dari setiap *cluster*. *Cluster 0* memiliki jumlah penjualan terbanyak pada produk “Golda Coffe 200ml” dengan 354 penjualan. *Cluster 1* memiliki jumlah penjualan terbanyak pada produk “Mie Sedap Baso Spesial” dengan 844 penjualan. *Cluster 2* memiliki jumlah penjualan terbanyak pada produk “Telor ½ kg” dengan 3556 penjualan.

### 3.4. Evaluasi Hasil

Metode *confusion matrix* digunakan sebagai evaluasi hasil dari model klusterisasi pada penelitian ini. *Confusion matrix* menghasilkan nilai *accuracy*, *recall*, *precision* dan *f1-score*. Hasil dari evaluasi ini dapat dilihat pada Gambar 5.

	precision	recall	f1-score	support
0	1.00	0.88	0.94	7345
1	0.00	0.00	0.00	5
2	0.00	0.00	0.00	53
3	0.00	0.00	0.00	0
4	0.00	0.00	0.00	0
5	0.00	0.00	0.00	0
6	0.00	0.00	0.00	0
7	0.00	0.00	0.00	0
8	0.00	0.00	0.00	0
accuracy			0.87	7403
macro avg	0.11	0.10	0.10	7403
weighted avg	0.99	0.87	0.93	7403

Gambar 5. Hasil *Confusion Matrix*

Berdasarkan Gambar 7, nilai akurasi yang didapatkan oleh model klusterisasi dengan *k-means* pada penelitian ini adalah 87%. Nilai akurasi ini masih lebih tinggi dari pada model klusterisasi yang dilakukan oleh penelitian [10] dan [13]. Nilai akurasi namun masih lebih rendah dibandingkan model klusterisasi yang dilakukan oleh penelitian [11] dan [12].

#### 4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, algoritma *k-means clustering* dapat digunakan untuk melakukan analisa data transaksi penjualan pada Mini Market Berkah Abadi Tegal. Algoritma ini berhasil menemukan tiga jenis *cluster* data penjualan yaitu produk dengan penjualan sedikit, sedang dan banyak yang masing-masing disimbolkan dengan *cluster 0*, *cluster 1* dan *cluster 2*. Jumlah produk dengan transaksi dan total penjualan terbanyak atau *cluster 2* sebanyak 57 produk dan untuk produk dengan transaksi dan total penjualan sedang atau *cluster 1* sebanyak 5 produk dan sisanya merupakan produk dengan total penjualan sedikit. Model klusterisasi dengan *k-means* juga menghasilkan akurasi sebesar 87%. Berdasarkan hasil ini pemilik Minimarket Berkah Abadi Tegal dapat terbantu dalam membuat keputusan pada manajemen stok barang terutama pada saat pandemi Covid-19 masih berlangsung. Kekurangan pada penelitian ini adalah analisa yang dilakukan masih terbatas pada data yang masih terbilang sedikit dan implementasinya terbatas pada data *console* dengan *python* dan belum diaplikasikan pada sistem dengan berbasis *user interface* seperti website. Rencana penelitian selanjutnya adalah melakukan analisa pada data transaksi dari minimarket lain dan harapannya juga dapat mengelola data dengan jumlah yang lebih besar. Kami juga akan menyempurnakannya dalam sistem informasi berbentuk website agar pemilik minimarket dengan mudah melakukan analisa data transaksi mereka.

#### Daftar Pustaka

- [1] I. Wahidah, R. Athallah, N. F. S. Hartono, M. C. A. Rafqie, and M. A. Septiadi, "Pandemik COVID-19: Analisis Perencanaan Pemerintah dan Masyarakat dalam Berbagai Upaya Pencegahan," *J. Manaj. dan Organ.*, vol. 11, no. 3, pp. 179–188, 2020, doi: 10.29244/jmo.v11i3.31695.
- [2] F. R. Yamali and R. N. Putri, "Dampak Covid-19 Terhadap Ekonomi Indonesia," *Ekon. J. Econ. Bus.*, vol. 4, no. 2, p. 384, 2020, doi: 10.33087/ekonomis.v4i2.179.
- [3] A. Amri, "Dampak Covid-19 Terhadap UMKM di Indonesia," *J. Brand*, vol. 2, no. 1, pp. 123–130, 2020.
- [4] Y. Mardi, "Data Mining: Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [5] M. F. Mulya, N. Rismawati, and A. R. Rizky, "Analisis Dan Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Penjualan Pada Kantin Universitas Tanri Abeng," *Fakt. Exacta*, vol. 12, no. 3, pp. 210–218, 2019, doi: 10.30998/faktorexacta.v12i3.4541.
- [6] D. Firdaus, "Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer," *J. Format*, vol. 6, no. 2, pp. 91–97, 2017.
- [7] M. H. Siregar, "Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan)," *J. Teknol. Dan Open Source*, vol. 1, no. 2, pp. 83–91, 2018, doi: 10.36378/jtos.v1i2.24.
- [8] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri)," *J. Sist.*, vol. 7, no. September, pp. 238–249, 2018.
- [9] W. Afifi, D. R. Nastiti, and Q. Aini, "Clustering K-Means Pada Data Ekspor (Studi Kasus: Pt. Gaikindo)," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 1, pp. 45–50, 2020, doi: 10.24176/simet.v11i1.3568.
- [10] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *INFORMAL Informatics J.*, vol. 5, no. 1, p. 10, 2020, doi: 10.19184/isj.v5i1.17071.
- [11] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [12] D. Abipraya, V. Christianti, and M. N. Jaya, "Implementasi Opinion Mining Untuk Provider Internet Menggunakan Metode Naive Bayes," *Ilmu Komput. dan Sist. Inf.*, vol. 9, no. 2, pp. 1–8, 2020.
- [13] A. U. Fitriyadi and A. Kurniawati, "Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional," *KILAT*, vol. 10, no. 1, pp. 157–168, 2021.
- [14] I. made B. Adnyana, "Penerapan Feature Selection untuk Prediksi Lama Studi Mahasiswa," *J. Sist. Dan Inform.*, vol. 13, pp. 72–76, 2019.
- [15] F. Nasari and C. J. M. Sianturi, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat," *CogITO Smart J.*, vol. 2, no. 2, p. 108, 2016, doi: 10.31154/cogito.v2i2.19.108-119.
- [16] L. A. A. R. Putri, "Seleksi Fitur Dalam Klasifikasi Genre Musik," *J. Ilm. Ilmu Komput. Udayana*, vol. 10, no. 1, pp. 19–26, 2017.
- [17] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.
- [18] E. Muningsih and S. Kiswati, "Sistem Aplikasi

- Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan,” *Joutica*, vol. 3, no. 1, p. 117, 2018, doi: 10.30736/jti.v3i1.196.
- [19] A. B. H. Kiat, Y. Azhar, and V. Rahmayanti, “Penerapan Metode K-Means Dengan Metode Elbow Untuk Segmentasi Pelanggan Menggunakan Model RFM(Recency, Frequency, & Monetary),” *J. Repos.*, vol. 2, no. 7, p. 945, 2020, doi: 10.22219/repositor.v2i7.973.
- [20] N. T. Hartanti, “Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 2, pp. 82–89, 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.
- [21] H. Priyatman, F. Sajid, and D. Haldivany, “Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan,” vol. 5, no. 1, pp. 62–66, 2019.
- [22] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [23] R. Muliono and Z. Sembiring, “Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen,” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 4, no. 2, pp. 2502–714, 2019.
- [24] I. Novandya, Adhika., Oktria, “Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi,” *J. Format*, vol. 6, no. 2, pp. 98–106, 2017, [Online]. Available: <https://www.neliti.com/publications/224664/penerapan-algoritma-klasifikasi-data-mining-c45-pada-dataset-cuaca-wilayah-bekas>.
- [25] R. Umar, I. Riadi, and Purwono, “Perbandingan Metode SVM, RF dan SGD untuk Penentuan Model Klasifikasi Kinerja Programmer pada Aktivitas Media Sosial,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 2, pp. 329–335, 2020.