

BAB II

TINJAUAN PUSTAKA

A. Penelitian Terdahulu

Beberapa penelitian terdahulu yang telah dilakukan oleh beberapa penulis:

1. (Tinuke Omolewa et al., 2019) dari *Department of Computer Science, University of Ilorin, Ilorin, Nigeria*, melakukan analisa data kinerja siswa menggunakan metode *K-Means Clustering* untuk menilai kinerja akademik siswa. Hasil penelitian ini menunjukkan bahwa nilai ujian siswa, kuis, dan tugas merupakan faktor utama yang dapat digunakan untuk memprediksi kinerja akademik siswa, diperoleh dua kluster dengan menggunakan metode siku untuk mengelompokkan semua siswa ke dalam kluster.
2. (Mathematics, 2021) dari *School of Business Management, Nanyang Polytechnic, Singapore* melakukan analisa data penjualan pada Toko Basket menggunakan metode *K-Means Clustering*. Melakukan segmentasi menjadi enam data kluster untuk menentukan strategi penjualan lebih baik dengan melakukan evaluasi dari beragam produk yang ada.
3. (Abid et al., 2022) menganalisa penjualan komoditas Toko Tani Indonesia untuk dapat mengelompokkan jenis barang yang laris terjual dan tidak sehingga barang di gudang tidak menumpuk. Algoritma yang digunakan dalam teknik data mining untuk klustering adalah *K-Means*. Penelitian ini sangat membantu mempermudah Toko Tani Indonesia Center untuk mengembangkan strategi persediaan stok barang.
4. (Indriyani & Irfiani, 2019) melakukan klustering data penjualan pada Toko Perlengkapan *Outdoor* menggunakan metode *K-Means*. Penerapan metode ini dalam mengelompokkan penjualan pada toko dapat memberikan rekomendasi barang laris, kurang laris, dan cukup laris. Hasil penelitian ini dijadikan rujukan untuk mengatur stok barang agar barang yang ingin dibeli pelanggan tidak kosong.
5. (Novalia & Ahmad, 2022) melakukan klasifikasi hasil penjualan minuman ringan pada koperasi berdasarkan jenis barang. Penelitian ini menggunakan

algoritma *K-Means Clustering*. Hasil dari penelitian ini adalah untuk mengelompokkan hasil penjualan ke dalam sebuah kluster sesuai dengan karakteristiknya.

6. (Aulia, 2021) melakukan klasterisasi penjualan pestisida menggunakan metode *K-Means Clustering* di Toko Juanda Tani. Pengelompokan ini bertujuan agar pihak toko dapat mengetahui penjualan pestisida jenis apa aja yang laku dan tidak laku selama satu musim agar tidak terjadi penumpukan barang di gudang.
7. (M. Y. Rizki et al., 2020) melakukan penelitian menggunakan metode *K-Means Clustering* untuk mengelompokkan jumlah penjualan ikan laut di TPI menurut wilayah. Penelitian ini diharapkan dapat menjadi informasi kepada pemerintah wilayah Indonesia tentang jumlah penjualan ikan laut di Indonesia dan menjadi acuan untuk menyeimbangkan penjualan ikan. Peneliti memanfaatkan *software RapidMiner* dalam pengujian penelitian ini. Hasil dari penelitian ini menghasilkan 2 Provinsi menduduki posisi kluster tingkat tinggi (C1) dan 18 provinsi menduduki kluster tingkat rendah (C2).
8. (Kurniawan & Enri, 2021) melakukan penelitian menggunakan algoritma *K-Means* untuk mengevaluasi kinerja *Assistant Sales Representative* (ASR) pada PT Pupuk Kujang Cikampek, Departemen Pemasaran dan Penjualan Ritel. Penelitian ini dilakukan pengelompokan dengan tiga kluster yaitu produktif, menengah produktif dan kurang produktif. Hasil dari penelitian ini didapatkan sebanyak 6 ASR masuk ke dalam kelompok kurang produktif, 22 ASR masuk ke dalam kelompok menengah produktif, dan 8 ASR masuk ke dalam kelompok produktif.
9. (Ramadhan et al., 2021) melakukan pengelompokan penjualan paket data menggunakan algoritma *Clustering K-Means*. Pada penelitian ini terdapat 40 data yang digunakan dalam uji coba, kemudian ditentukan tiga kluster yang digunakan sebagai acuan dalam iterasi awal pada metode yang digunakan. Kluster ini dibagi menjadi tiga yaitu paket banyak, sedang, dan sedikit. Hasil dari uji coba yang dilakukan diperoleh hasil optimal pada

iterasi ketiga dengan hasil akhir yang menunjukkan bahwa hasil terendah yaitu C1 (*Cluster 1*), hasil sedang yaitu C3 (*Cluster 3*), dan hasil tertinggi yaitu C2 (*Cluster 2*).

10. (Febianda et al., 2020) melakukan pengelompokan penjualan *Rattan Furniture* pada PT. Hyma Indotraco berbasis Algoritma *K-Means Clustering*. Penelitian ini bertujuan untuk mengetahui pola pengelompokan data penjualan menggunakan *K-Means Clustering* untuk mendapatkan hasil pengelompokan penjualan *rattan furniture* yang paling banyak dipesan, sering dipesan, dan jarang dipesan. Dalam implementasinya menggunakan *tools RapidMiner*. Penelitian ini menghasilkan kluster yang jarang dipesan sebanyak 678 produk, kluster yang sering dipesan sebanyak 14 produk, dan kluster yang paling banyak dipesan sebanyak 15 produk.
11. (Pii et al., 2023) melakukan penelitian pada penjualan produk pakaian Dameyra Fashion menggunakan metode *K-Means Clustering*. Permasalahan yang dihadapi adalah masih belum bisa mengelola persediaan barang, sehingga diterapkan metode ini untuk mengelompokkan data penjualan toko menjadi 2 kluster. Hasil yang didapat yaitu penjualan yang laris ada 289 item dan tidak laris 9 item. Dengan ini dapat diketahui rincian produk yang tidak laris, sehingga ke depannya bisa memaksimalkan persediaan stok untuk mengoptimalkan penjualan.
12. (Puji Rahayu et al., 2020) melakukan analisa penjualan rotan menggunakan metode *K-Means Clustering*. Banyaknya jumlah data akan lebih sulit untuk dianalisa, sehingga analisa menggunakan metode ini dirasa cocok digunakan untuk melihat pola penjualan produk. Hasil dari penelitian ini menghasilkan 3 kluster, yaitu tidak laku, sedang laku, dan laris. Kelompok tidak laku sejumlah 6 data, sedang laku sejumlah 114 data dan laris sejumlah 9 data.
13. (Muni, 2020) melakukan analisis algoritma *K-Means Clustering* untuk menentukan strategi promosi penjualan sepeda motor di PT. Alfa Scorpii. Penggunaan data tidak maksimal mengakibatkan laporan penjualan yang digunakan hanya menjadi sebuah arsip. Pengolahan data menggunakan

metode ini bisa membantu strategi pemasaran yang tepat karena dapat mengelompokkan barang untuk memaksimalkan penjualan. Hasil yang didapat yaitu terdapat 2 kluster, yang laris dan tidak laris. Kelompok laris terdapat 36 item dan tidak laris terdapat 21 item.

14. (Rasyid et al., 2022) melakukan pengelompokan penjualan produk pada Online Shop Toko Gizi menggunakan metode *K-Means Clustering*. Toko Gizi memiliki riwayat transaksi yang berisi banyaknya penjualan dan transaksi, dimana nantinya data tersebut dapat digunakan dalam *Data Mining* untuk melihat perputaran barang dari yang kurang diminati hingga sangat diminati, yang harapannya dapat digunakan untuk Toko Gizi dalam mempersiapkan stok barang yang disesuaikan dari penjualan dalam rentang waktu tertentu. Algoritma *K-Means* ini cocok digunakan untuk mengatur riwayat data pembelian dan penjualan karena implementasinya yang efisien, tidak perlu memakan banyak waktu serta penerapannya cukup mudah dibandingkan teknik *clustering* lainnya. Dari penelitian yang sudah dilakukan telah didapatkan hasil dari metode *Clustering* algoritma *K-Means* yang menunjukkan bahwa dari 41 jenis data penjualan dan transaksi yang dilakukan dari bulan April hingga Mei 2022 menghasilkan 3 cluster yaitu sangat diminati, cukup diminati dan kurang diminati. Dimana dari beberapa cluster tersebut menunjukkan bahwa cluster pertama terdapat 5 barang, cluster kedua terdapat 5 barang dan cluster ketiga terdapat 1 barang.
15. (Syahputra, 2022) melakukan penelitian tingkat penjualan menu *Food and Beverage* menggunakan algoritma *K-Means Clustering* untuk menentukan efektifitas penjualan dan bahan baku. Sumber data penelitian ini dari histori transaksi penjualan menu selama 1 tahun, kemudian dianalisis dengan algoritma *K-Means*. Informasi yang ditemukan berupa menu F&B yang populer dan tingkat penjualan menu. Tujuan penelitian ini untuk melakukan pengelompokan pada data tingkat penjualan menu (*Food and Beverage*). Hasil penelitian menunjukkan bahwa metode *Clustering* dengan algoritma *K-Means* memberikan keluaran yang cukup signifikan dalam mengelompokkan

data penjualan. Kontribusi penelitian memberikan pengetahuan dalam bentuk sebuah informasi dalam melakukan manajemen data penjualan.

16. (Astuti et al., 2022) melakukan penelitian pada penjualan produk digital konter di Leppangeng Cell menggunakan metode *K-Means Clustering*. Konter Leppangeng Cell merupakan sebuah konter yang menjual produk digital berupa Pulsa, Paket Data, Token Listrik PLN, Pembayaran Listrik, Pembayaran BPJS, Pembayaran PDAM, Pembayaran Spidy atau Indihome, Pembayaran Telephone Rumah, Pembayaran Pasca Bayar, Penjualan Voucher Game, Pembayaran TV Kabel, Pembayaran Angsuran Kredit, Pembayaran Tiket Kereta, Pembayaran PGN/Gas, Pembayaran Shopee dan Jasa Transfer tunai dan Top Up. Namun, terjadi penjualan yang tidak seimbang dan data riwayat penjualan tidak digunakan dengan baik, hanya untuk arsip toko saja. Melakukan penerapan *data mining* dengan metode ini bertujuan untuk mengetahui penjualan produk digital yang laris dan yang tidak melalui pengelompokan data menjadi 3 kluster dari data riwayat penjualan. Hasilnya yaitu didapatkan 114 produk laku, 5 produk kurang laku, dan 14 produk tidak laku, dimana hasil ini akan digunakan sebagai acuan untuk proses penyetokan barang selanjutnya.
17. (Rahayuni et al., 2022) melakukan klusterisasi penjualan material menggunakan metode *K-Means*. Hasil yang diperoleh telah diimplementasikan ke dalam *RapidMiner* dan memiliki tiga kluster, yaitu kluster tidak laku, kluster laku, dan kluster sangat laku. Dengan kluster tidak laku berwarna merah 2 item, kluster laku berwarna hijau 28 item, dan kluster sangat laku berwarna biru 30 item. Hasil penelitian ini bisa menjadi masukan kepada Toko Family Gypsum Jl. H. Ulakma Sinaga, Rambung Merah yang menjadi perhatian lebih pada tiap-tiap penjualannya.

B. Landasan Teori

1. Penjualan

Penjualan adalah kegiatan menjual tenaga kerja atau produk dengan harapan mendapatkan keuntungan dari pertukaran yang dilakukan. Mengingat pengertian di atas, dapat dikatakan bahwa transaksi adalah gerakan untuk menjual tenaga kerja dan produk dan peristiwa pertukaran untuk memindahkan tanggung jawab atas tenaga kerja dan produk yang diusulkan kepada pembeli di mana pembeli dapat menunjukkan sejumlah uang tunai sebagai alat penaksir barang atau administrasi dengan biaya yang disetujui (Febianda et al., 2020).

2. Kosmetik

Kosmetik atau produk perawatan kecantikan adalah bahan perawatan yang digunakan untuk memperbaiki penampilan atau tubuh manusia. Kosmetik biasanya dibuat dari kombinasi bahan kimia, bahan alami, atau bahan sintetis. Produk perawatan ini digunakan untuk membuat tubuh terlihat atau berbau lebih baik. Proses pengaplikasian kosmetik disebut sebagai "tata rias" atau nama lain untuk itu (Latief & Ayustira, 2020).

3. Data Mining

Data mining adalah proses pengumpulan data historis untuk menemukan hubungan dan keteraturan dalam kumpulan data yang besar. Proses menganalisis data dengan tujuan membangun hubungan yang jelas dan sebelumnya tidak dikenali antara data dan sumber lain sehingga dapat dimanfaatkan dan dipahami disebut sebagai *data mining* atau penambangan data. *Data mining* adalah proses menemukan hubungan pola, dan tren dalam data yang tersimpan dalam memori dalam jumlah besar dengan menggunakan alat untuk pengenalan pola seperti statistik dan matematika. (Pii et al., 2023).

Menurut (Indriyani & Irfiani, 2019), *data mining* digunakan untuk menangani informasi yang sangat besar dalam kumpulan data untuk menciptakan data baru yang berharga bagi prosedur bisnis. Penambangan

deskriptif dan prediktif adalah dua jenis utama dari penambangan data, dengan penambangan deskriptif digunakan untuk mengidentifikasi karakteristik data dan penambangan prediktif digunakan untuk mengidentifikasi pola data. Tujuan dari *data mining* adalah untuk memberikan Toko Kosmetik Azakayra solusi nyata untuk menentukan produk paling populer yang diminati pelanggan.

4. Algoritma *K-Means Clustering*

Algoritma *K-Means Clustering* adalah teknik analisis kelompok yang digunakan dalam statistik dan pembelajaran mesin yang membagi N pengamatan menjadi K kelompok (*cluster*), dengan setiap pengamatan menjadi bagian dari kelompok dengan rata-rata terdekat. Hal ini mirip dengan algoritma *Expectation-Maximization* yang digunakan dalam *Gaussian Mixture*, yang mencoba menemukan pusat kelompok dalam data sebanyak mungkin (Sallaby et al., 2022).

Metode *K-Means* melibatkan langkah-langkah berikut untuk pengelompokan:

- 1) Dari kumpulan sampel data berikut, tentukan jumlah *cluster* (k) untuk jumlah cluster. Perhitungan dimulai dengan angka K , yang merepresentasikan jumlah kelompok cluster ($K = 3$) bisa dilihat pada tabel berikut:

Data ke	Nama Produk	X	Y
1	Wardah Lightening Powder Foundation Light Feel	13	624.000
2	Wardah Lightening Day Cream 17ml	17	459.000
3	Wardah Lightening Night Cream 17ml	44	1.254.000
4	Wardah Lightening Night Cream 30gr	6	228.000

5	Wardah Lightening Face Toner 125ml	37	962.000
---	---------------------------------------	----	---------

Keterangan :

X = Total Penjualan Produk

Y = Total Income

- 2) Tentukan *centroid* atau pusat cluster secara acak. Untuk penentuan titik pusat cluster bisa dilihat pada tabel di bawah:

Cluster	X	Y
C1	6	228.000
C2	17	459.000
C3	44	1.254.000

- 3) Tentukan *centroid* terdekat untuk setiap objek data. Adapun persamaan yang sering digunakan dalam pemecahan masalah dalam menentukan jarak terdekat adalah menggunakan persamaan *Euclidean Distance*. Untuk rumus perhitungannya bisa dilihat di bawah ini.

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

$D(i, j)$ = Jarak data ke i ke pusat cluster j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

Untuk contoh perhitungannya bisa dilihat pada perhitungan sebagai berikut:

- a. Perhitungan data ke-1 terhadap cluster 1, 2, 3

Jarak data 1 ke cluster 1 :

$$(13, 624.000) = \sqrt{(13 - 6)^2 + (624.000 - 228.000)^2} = 396.02$$

Jarak data 1 ke cluster 2:

$$(13, 624.000) = \sqrt{(13 - 17)^2 + (624.000 - 459.000)^2} = 165.01$$

Jarak data 1 ke cluster 3:

$$(13, 624.000) = \sqrt{(13 - 44)^2 + (624.000 - 1.254.000)^2} = 630225.004$$

b. Perhitungan data ke-2 terhadap cluster 1, 2, 3

Jarak data 2 ke cluster 1 :

$$(17, 459.000) = \sqrt{(17 - 6)^2 + (459.000 - 228.000)^2} = 231.24$$

Jarak data 2 ke cluster 2 :

$$(17, 459.000) = \sqrt{(17 - 17)^2 + (459.000 - 459.000)^2} = 0$$

Jarak data 2 ke cluster 3 :

$$(17, 459.000) = \sqrt{(17 - 44)^2 + (459.000 - 1.254.000)^2} = 795009.995$$

c. Perhitungan data ke-3 terhadap cluster 1, 2, 3

Jarak data 3 ke cluster 1 :

$$(44, 1.254.000) = \sqrt{(44 - 6)^2 + (1.254.000 - 228.000)^2} = 1,026,902.03$$

Jarak data 3 ke cluster 2 :

$$(44, 1.254.000) = \sqrt{(44 - 17)^2 + (1.254.000 - 459.000)^2} = 795,015.06$$

Jarak data 3 ke cluster 3 :

$$(44, 1.254.000) = \sqrt{(44 - 44)^2 + (1.254.000 - 1.254.000)^2} = 0$$

d. Perhitungan data ke-4 terhadap cluster 1, 2, 3

Jarak data 4 ke cluster 1 :

$$(6, 228.000) = \sqrt{(6 - 6)^2 + (228.000 - 228.000)^2} = 0$$

Jarak data 4 ke cluster 2 :

$$(6, 228.000) = \sqrt{(6 - 17)^2 + (228.000 - 454.000)^2} = 226.26$$

Jarak data 4 ke cluster 3 :

$$(6, 228.000) = \sqrt{(6 - 44)^2 + (228.000 - 1.254.000)^2} = 1,026,902.03$$

e. Perhitungan data ke-5 terhadap cluster 1, 2, 3

Jarak data 5 ke cluster 1 :

$$(37, 962.000) = \sqrt{(37 - 6)^2 + (962.000 - 228.000)^2} = 734,018.86$$

Jarak data 5 ke cluster 2 :

$$(37, 962.000) = \sqrt{(37 - 17)^2 + (962.000 - 459.000)^2} = 503,000.10$$

Jarak data 5 ke cluster 3 :

$$(37, 962.000) = \sqrt{(37 - 44)^2 + (962.000 - 1.254.000)^2} = 292,000.01$$

- 4) Lakukan pengelompokan data berdasarkan kedekatan hasil perhitungan sebelumnya dengan *centroid*. Hasil ini akan menunjukkan tipe cluster mana pada setiap data yang bisa dilihat pada tabel berikut:

Data ke-	X	Y	C1	C2	C3	Cluster
1	13	624.000	396.02	165.01	630225.009	2
2	17	459.000	231.29	0	759009.995	2
3	44	1.254.000	1,026,902.03	795,015.06	0	3
4	6	228.000	0	226.26	1,026,902.03	1

5	37	962.000	734,018.06	503,000.10	292,000.01	3
---	----	---------	------------	------------	------------	---

Pada tabel tersebut, penentuan clustering dilihat dapat dilihat dari kedekatan jarak data terhadap titik pusat yang ditentukan, hasilnya ditandai dengan blok warna kuning.

- 5) Selanjutnya adalah melakukan *update* data dengan memperbarui nilai *centroid* melalui beberapa kali iterasi. Jika hasilnya sudah tetap dan tidak berubah lagi dengan hasil pada iterasi sebelumnya, maka proses iterasi dihentikan dan ditemukanlah hasil clustering yang sudah optimal.

Cluster baru tersebut didapat dari rumus = nilai hasil / banyak hasil

Cluster 1

Nilai X = $6/1 = 6$

Nilai Y = $228.000/1 = 228.000$

Cluster 2

Nilai X = $(13+17)/2 = 15$

Nilai Y = $(624.000+459.000)/2 = 541.500$

Cluster 3

Nilai X = $(44+37)/2 = 40.5$

Nilai Y = $(1.254.000+962.000)/2 = 1.108.000$

Didapat nilai *centroid* baru diantaranya:

Cluster	X	Y
C1	6	228.000
C2	15	541.500
C3	40.5	1.108.000

- 6) Setelah mendapat nilai *centroid* baru, langkah selanjutnya adalah melanjutkan perhitungan seperti pada langkah ke 3 dan 4 dengan iterasi ke-2 dan seterusnya hingga hasil cluster selalu tetap.

Perhitungan iterasi ke-2:

1) Perhitungan data ke-1 terhadap cluster 1, 2, 3

Jarak data 1 ke cluster 1 :

$$(13, 624.000) = \sqrt{(13 - 6)^2 + (624.000 - 228.000)^2} = 396.02$$

Jarak data 1 ke cluster 2:

$$(13, 624.000) = \sqrt{(13 - 15)^2 + (624.000 - 541.500)^2} = 82,499.84$$

Jarak data 1 ke cluster 3:

$$(13, 624.000) = \sqrt{(13 - 40.5)^2 + (624.000 - 1.108.000)^2} = 484,149.39$$

b. Perhitungan data ke-2 terhadap cluster 1, 2, 3

Jarak data 2 ke cluster 1 :

$$(17, 459.000) = \sqrt{(17 - 6)^2 + (459.000 - 228.000)^2} = 231.24$$

Jarak data 2 ke cluster 2 :

$$(17, 459.000) = \sqrt{(17 - 15)^2 + (459.000 - 541.500)^2} = 82,499.84$$

Jarak data 2 ke cluster 3 :

$$(17, 459.000) = \sqrt{(17 - 40.5)^2 + (459.000 - 1.108.000)^2} = 648,927.93$$

c. Perhitungan data ke-3 terhadap cluster 1, 2, 3

Jarak data 3 ke cluster 1 :

$$(44, 1.254.000) = \sqrt{(44 - 6)^2 + (1.254.000 - 228.000)^2} = 1,026,902.03$$

Jarak data 3 ke cluster 2 :

$$(44, 1.254.000) = \sqrt{(44 - 15)^2 + (1.254.000 - 541.500)^2} = 711,912.36$$

Jarak data 3 ke cluster 3 :

$$(44, 1.254.000) = \sqrt{(44 - 40.5)^2 + (1.254.000 - 1.108.000)^2} = 146,003.34$$

d. Perhitungan data ke-4 terhadap cluster 1, 2, 3

Jarak data 4 ke cluster 1 :

$$(6, 228.000) = \sqrt{(6 - 6)^2 + (228.000 - 228.000)^2} = 0$$

Jarak data 4 ke cluster 2 :

$$(6, 228.000) = \sqrt{(6 - 15)^2 + (228.000 - 541.500)^2} = 314,090.34$$

Jarak data 4 ke cluster 3 :

$$(6, 228.000) = \sqrt{(6 - 40.5)^2 + (228.000 - 1.108.000)^2} = 880,003.53$$

e. Perhitungan data ke-5 terhadap cluster 1, 2, 3

Jarak data 5 ke cluster 1 :

$$(37, 962.000) = \sqrt{(37 - 6)^2 + (962.000 - 228.000)^2} = 734,018.86$$

Jarak data 5 ke cluster 2 :

$$(37, 962.000) = \sqrt{(37 - 15)^2 + (962.000 - 541.500)^2} = 420,504.89$$

Jarak data 5 ke cluster 3 :

$$(37, 962.000) = \sqrt{(37 - 40.5)^2 + (962.000 - 1.108.000)^2} = 146,003.34$$

Setelah melakukan perhitungan iterasi ke-2, bisa dilihat hasil clusteringnya pada tabel berikut:

Data ke-	X	Y	C1	C2	C3	Cluster
1	13	624.000	396.02	82,499.84	484,149.39	2
2	17	459.000	231.29	82,499.84	648,927.93	2
3	44	1.254.000	1,026,902.03	711,912.36	146,003.34	3
4	6	228.000	0	314,090.34	880,003.53	1
5	37	962.000	734,018.86	420,504.89	146,003.34	3

Hasil dari tahapan yang pertama dan kedua tidak berubah, maka hasil sudah sesuai dengan pengelompokan kluster. Berikut adalah hasil dari pengelompokan tersebut:

Data ke	Nama Produk	X	Y	Cluster
1	Wardah Lightening Powder Foundation Light Feel	13	624.000	2
2	Wardah Lightening Day Cream 17ml	17	459.000	2
3	Wardah Lightening Night Cream 17ml	44	1.254.000	3
4	Wardah Lightening Night Cream 30gr	6	228.000	1
5	Wardah Lightening Face Toner 125ml	37	962.000	3

5. Evaluasi *Sum of Square Error (SSE)*

Sum of Square Error (SSE) merupakan hasil penjumlahan dari seluruh jarak masing-masing data dengan titik pusat clusternya. Semakin kecil nilai SSE yang didapat, semakin seragam data yang ada didalam masing-masing cluster, semakin baik cluster yang dihasilkan. Jumlah cluster k yang dihasilkan dari pengujian dengan *K-Means*, dievaluasi menggunakan teknik SSE. Semakin banyak jumlah k yang diujikan, maka nilai SSE semakin kecil. SSE merupakan cara dalam melakukan validasi cluster melalui jumlah kuadrat setiap anggota cluster menuju pusatnya. Semakin jauh jarak yang membentuk titik siku, maka jumlah cluster tersebut menjadi yang paling optimal. Rumus SSE adalah sebagai berikut (Jollyta et al., 2019).

$$SSE = \sum_{K=1}^K \sum_{x_i \in S_K} \|x_i - c_k\|_2^2$$