

Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means

Genta Triyandana^{1*}, Lala Aprianti Putri^{2**}, Yuyun Umaidah^{3**}

* Teknik Informatika, Universitas Singaperbangsa Karawang

genta.triyandana18117@student.unsika.ac.id¹, lala.aprianti18051@student.unsika.ac.id², yuyun.umaidah@staff.unsika.ac.id³

Article Info

Article history:

Received 2022-01-07

Revised 2022-04-28

Accepted 2022-05-9

Keyword:

Data Mining,
Clustering,
K-Means,
Tingkat Penjualan

ABSTRACT

Data mining can be used to find solutions in making sales decisions to increase sales. Sales data storage stores many sales transaction records, where each document provides products purchased by customers in each sales transaction. A problem began to arise with an excess stockpiling of materials. The number of fluctuating sales causes the stock of available materials to be unstable and can directly impact consumers. Mistakes in predicting sales caused the coffee shop to buy large quantities of material stock, which were not widely used or sold out, so the supply of these materials swelled in the warehouse. One way to be implemented is by applying data mining because there are ways and methods to meet needs, one of which is the need for extensive information, then the information that we can use to determine quality in determining a decision. Therefore, it is hoped that this research can help Dpom Coffee minimize material stock inventory management cases such as shortages and excesses and make policies to increase sales by grouping menus based on sales levels using the K-means algorithm. Based on the results of processing the sales dataset at Dpom Coffee, it produces 3 clusters, namely Cluster 1 with eight menus with low sales levels, cluster 2 with 40 menus with moderate sales levels, and cluster 3 with seven menus with high sales levels. The accuracy or performance of the k-means algorithm results in a Davies Bouldin index value of 0.457.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Dpom Coffee merupakan suatu UMKM yang bergerak di bidang kuliner. Dpom coffee telah berdiri sejak tahun 2020 yang memusatkan penjualan dengan berbagai macam menu minuman seperti kopi dan the serta makanan. Berkembangnya teknologi saat ini, kita dapat memanfaatkan untuk mempermudah dan memaksimalkan pekerjaan kita sehari – hari, baik dari aspek apapun. Dalam hal bisnis pun, transaksi yang telah terjadi dapat disimpan dalam bentuk digital pada sebuah komputer. Proses transaksi pun mengalami kemajuan dengan berbagai cara dan metode yang dilakukan untuk meningkatkan tingkat penjualansuatu produk. Maka dari itu diperlukan suatu analisa dalam sebuah bisnis untuk mempertahankan bahkan meningkatkan tingkat penjualan. Salah satu metode analisa yang digunakan yaitu penerapan Data mining dengan menggunakan pengelompokkan atau clustering. Data mining dapat

dimanfaatkan untuk mendapatkan solusi dalam suatu pengambilan keputusan dalam penjualan demi meningkatkan penjualan. Penyimpanan data penjualan menyimpan jumlah record transaksi penjualan yang besar, dimana setiap record memberikan produk yang dibeli oleh pelanggan dalam setiap transaksi penjualan.

Suatu masalah mulai muncul, dimana adanya penimbunan stok bahan yang berlebih. Jumlah penjualan yang fluaktif menyebabkan stok bahan yang tersedia tidak stabil dan dapat berdampak langsung ke konsumen. Jika ketersediaan bahan yang tidak dikelola dengan baik juga berdampak pada coffeeshop tersebut, dimana jika bahan habis saat permintaan konsumen tinggi, maka akan menyebabkan permintaan produk atau menu harus diganti atau dibatalkan, dimana hal tersebut dapat berdampak buruk pada suatu coffeeshop tersebut.

Kesalahan dalam memprediksi penjualan menjadi penyebab *coffeeshop* tersebut membeli stok bahan dalam

jumlah besar yang pada akhirnya tidak banyak digunakan atau tidak habis terjual sehingga stok bahan tersebut membengkak di gudang. Dengan demikian, penimbunan dapat mengakibatkan coffeeshop tersebut mengalami kerugian dikarenakan biaya yang harus keluar untuk proses penyimpanan bahan sehingga pimpinan kesulitan untuk mengetahui produk atau menu mana yang lebih diminati dan banyak terjual. Harga dapat menjadi penentu dalam keberhasilan suatu usaha [1], karena harga menentukan berapa banyak keuntungan yang akan diperoleh dari penjualan tersebut.

Salah satu cara yang dapat diimplementasikan yaitu dengan menerapkan penggunaan *Data mining*, karena terdapat cara dan metode untuk memenuhi kebutuhan, salah satunya yaitu kebutuhan informasi yang luas, lalu dari informasi yang telah kita dapat jika kita manfaatkan untuk menentukan suatu kualitas dalam menentukan suatu keputusan. *Data mining* dapat dimanfaatkan oleh pebisnis dalam pengambilan keputusan secara cepat dan juga tepat [2] Hasil dari *data mining* dapat dimanfaatkan dalam membantu pengambilan keputusan di masa mendatang [3] [4]

Penelitian terkait *data mining* dalam pengelompokkan dengan metode *k-means* pernah dilakukan oleh [3] menghasilkan bahwa dengan memanfaatkan metode *data mining* menggunakan *K-means* dapat sangat membantu dan memudahkan toko Helai dalam mengembangkan strategi persediaan stok baju. Dalam penelitian [5], yang menghasilkan model pengelompokkan dengan *k-means* menghasilkan akurasi sebesar 87%. Penelitian lain yang dilakukan [6] menghasilkan nilai akurasi model pengelompokkan sebesar 81%. Penelitian yang dilakukan oleh Handoko pada model pengelompokkan penjualan produk Telkomsel menggunakan *k-means* menghasilkan akurasi sebesar 99%.

Maka dari itu diharapkan penelitian ini dapat membantu *Dpom Coffee* untuk meminimalisir kasus manajemen persediaan stok bahan seperti kekurangan maupun kelebihan dan membuat kebijakan dengan tujuan meningkatkan penjualan. Penelitian lain [7] juga menjelaskan, memahami keinginan serta kebutuhan konsumen merupakan hal yang penting untuk mencapai keberhasilan dalam memasarkan suatu produk. Selanjutnya penelitian [8] menjelaskan, pengaruh pertumbuhan penjualan mencerminkan keberhasilan investasi periode masa lalu serta dapat dimanfaatkan sebagai indikator permintaan konsumen.

Berdasarkan yang sudah dijelaskan diatas, maka pada penelitian kali ini akan dilakukan pengelompokkan menu pada *Dpom coffee* berdasarkan tingkat penjualan dengan menggunakan algoritma *K-means*. Diharapkan pemilik *coffeeshop* dapat melakukan pemasaran dengan strategi yang tepat dalam melayani kebutuhan konsumen.

II. TINJAUAN PUSTAKA

Data mining dapat disebut sebagai suatu proses pencarian atau penambangan demi mendapatkan informasi dari sekumpulan data yang berjumlah besar (*Big data*) [9]. Hasil dari penambangan informasi tersebut kemudian digunakan untuk menguraikan pengetahuan yang bertujuan untuk menentukan sebuah keputusan dalam strategi bisnis [10]. Menurut [11], *Data mining* dikelompokkan menjadi beberapa bagian, yaitu Deskripsi, Klasifikasi [12], Estimasi, Prediksi, Pengelompokkan, dan Asosiasi.

Menurut [13], *Clustering* atau Pengelompokkan adalah metode statistik yang digunakan untuk mengelompokkan banyak data atau objek kedalam kelompok – kelompok berdasarkan karakteristik yang dimiliki objek atau data tersebut. *Clustering* merupakan metode yang bersifat *unsupervised* atau tanpa arahan dimana karakteristik tiap kluster tidak ditentukan sebelumnya, melainkan menurut kemiripan atribut – atribut dari suatu kelompok.

Pada penelitian sebelumnya yang dilakukan oleh [14] dengan menerapkan *data mining* dengan metode *k-means* pada toko sepatu merupakan sistem yang tepat bagi pemilik toko sepatu dalam menentukan penjualan produk yang diminati dan kurang diminati.

K-Means merupakan salah satu algoritma *clustering* yang digunakan dalam proses *Data mining*. *K-Means* bekerja dengan cara melakukan pengelompokkan secara partisi yang memisahkan data kedalam kelompok –kelompok tertentu dengan meminimalkan rata – rata jarak setiap data klasternya. Metode ini merupakan metode pengklasteran atau pengelompokkan yang paling terkenal dan banyak digunakan karena sederhana, mudah diimplementasikan, serta mampu mengelompokkan data yang besar dan kompleksitas waktunya linear $O(nKT)$ dengan n merupakan jumlah dokumen, K yaitu jumlah klaster dan T adalah jumlah iterasi. Metode ini dikembangkan oleh Mc Queen pada tahun 1967 [15]. Perbandingan algoritma *K-means* dengan *K-Medoids* [16] dalam melakukan pengelompokkan data transaksi bongkar muat, yang menghasilkan perbedaan yang tidak begitu jauh atau signifikan, perbedaan hanya terdapat diwaktu komputasi *k-means* lebih cepat dengan waktu 1 detik sedangkan *k-medoids* membutuhkan waktu sebanyak 1 menit 38 detik, yang berarti semakin banyak iterasi maka proses pengolahan data akan semakin lama.

Dalam menghitung jarak antara data dan centroid akan menggunakan persamaan *Euclidean Distance*. *Euclidean Distance* adalah suatu metode untuk menghitung jarak dari 2 buah titik dalam *Euclidian space* [17]. Berikut persamaan atau rumus dari *Euclidian Distance*:

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

Keterangan:

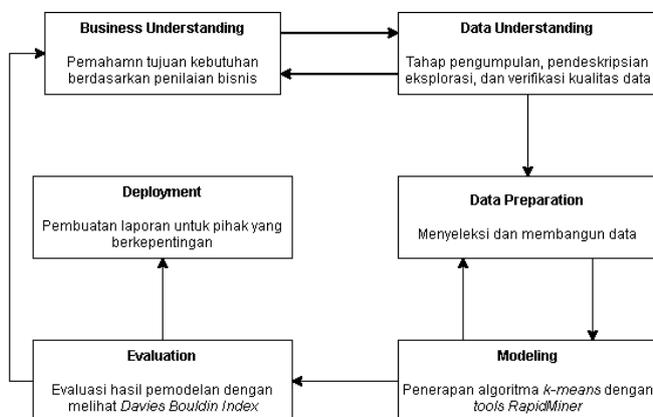
d = jarak antara x dan y

x = data pusat *cluster*
 y = data atribut
 i = setiap data
 n = banyak/jumlah data
 x_i = data pada pusat *cluster* ke i
 y_i = data pada setiap data ke i

Adapun dalam algoritma *k-means clustering* terdapat salah satu uji validitas internal pada metode yaitu Davies-Bouldin Index (DBI) [18] [19]. DBI berlandaskan dari nilai kohesi dan separasi. Dimana kohesi adalah jumlah dari kedekatan data terhadap *centroid* dari *cluster* yang diikuti. Sedangkan separasi merupakan jarak antar *centroid* dari *cluster* tersebut [6]. Dalam melakukan implementasi atau pengolahan data, akan menggunakan *tools Microsoft excel* dan *RapidMiner*. *Microsoft excel* adalah program aplikasi atau *tools* untuk mengolah angka yang dibuat atau dikeluarkan oleh *Microsoft Corp*. Sedangkan *RapidMiner* adalah aplikasi atau perangkat lunak yang dapat digunakan untuk pengolahan data berdasarkan prinsip serta algoritma *Data mining*.

III. METODE PENELITIAN

Dalam penelitian ini akan menggunakan metode penelitian Cross Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM adalah suatu standar dalam proses data mining yang berfungsi untuk melakukan tahapan analisis dari suatu industri sebagai dasar strategi pemecahan masalah bisnis atau suatu penelitian [20].



Gambar 1. Flowchart Metode Penelitian

Berikut penjelasan dari setiap tahapan yang ada di atas.

A. Business Understanding

Pada awal tahapan ini akan berfokus mengenai pemahaman tujuan berdasarkan penilaian bisnis dimana dari pemahaman tersebut akan diubah menjadi sebuah rencana awal data mining yang akan dirancang demi mencapai tujuan.

B. Data Understanding

Pada tahap selanjutnya akan dilakukan kegiatan berupa pengumpulan, pendeskripsian, dan eksplorasi data mana yang akan bermanfaat bagi Dinas Kesehatan Kabupaten Karawang, kemudian dari data tersebut diidentifikasi dan dipelajari data yang mana yang akan digunakan dalam penelitian. Data didapatkan dari Dpom Coffee terkait data penjualan dari bulan Januari – November 2021.

C. Data Preparation

Pada tahap ini akan dilakukan terkait persiapan terhadap data yang telah didapatkan sebelumnya. Persiapan yang dilakukan yaitu seperti data selection, data preprocessing, dan Transformation. Tahap ini dilakukan guna membangun data mentah awal menjadi dataset akhir yang selanjutnya akan digunakan pada tahap pemodelan.

D. Modeling

Pada tahap ini akan dilakukan proses dalam menentukan dan penerapan teknik atau metode data mining terhadap dataset akhir yang telah disiapkan untuk mengelompokkan menu makanan dan minuman pada Dpom Coffee. Pada penelitian ini akan menggunakan pemodelan algoritma K-means Clustering.

E. Evaluation

Pada tahap ini akan dilakukan analisa atau pengukuran ketepatan terhadap pemodelan yang telah dilakukan. Evaluasi dilakukan dengan menerapkan metode Davies Bouldin Index (DBI), dimana metode ini merupakan metode pengujian kualitas berdasarkan klaster yang dihasilkan. Evaluasi ditujukan untuk mengetahui pemodelan yang dilakukan apakah sudah tepat dan sesuai diterapkan pada kasus penelitian ini serta sudah sesuai rencana awal penelitian. Selanjutnya dari hasil evaluasi tersebut adalah menentukan langkah berikutnya apakah bisa dilanjutkan atau diulang dari awal karena tidak sesuai dengan rencana awal penelitian.

F. Deployment

Tahap selanjutnya yaitu tahap penyebaran hasil dari penelitian yang sudah dilakukan dijadikan sebagai laporan atau presentasi dari pengetahuan yang telah didapat berdasarkan pemodelan dan evaluasi pada proses data mining.

IV. HASIL DAN PEMBAHASAN

Dataset yang akan digunakan pada penelitian ini adalah data penjualan di Dpom Coffee pada tahun 2021. Data tersebut terdiri dari 55 data menu dengan 2 variabel. Berikut data penjualan dari Dpom Coffee yang ditampilkan pada Tabel 1.

TABEL 1.
DATASET PENJUALAN DPOM COFFEE

Product	Harga	Total Penjualan
V60	15000	216
Tubruk	10000	27
Vietnam Drip Hot	13000	257
Vietnam Drip Ice	15000	100
Espresso	10000	178
Americano Hot	13000	139
Americano Ice	15000	57
Cappucino Hot	15000	90
Cappucino Ice	16000	389
Cafe Latte Hot	15000	332
Cafe Latte Ice	16000	371
Caramel latte Hot	15000	60
Caramel latte Ice	16000	344
Hazelnut Latte Hot	15000	54
Hazelnut Latte Ice	16000	320
Vanilla Latte Hot	15000	12
Vanilla Latte ice	16000	467
KOPASUS	15000	774
Air Mineral 350ml	5000	1020
Moccacino Hot	14000	40
Moccacino Ice	15000	53
Matcha Hot	14000	10
Matcha Ice	15000	769
Coklat Hot	14000	65
Coklat Ice	15000	434
Butter Bir	15000	15
Kopi Bir	15000	246
Lemon tea	10000	964
Geraldine	20000	102
Japanese Ice	15000	90
Caramel Machiato Ice	23000	47
Jasmine Latte	18000	75
Rajampat	20000	274
Blue Bottels	20000	40
Susu Putih	10000	137
Lychee Tea	17000	241
Hazelnut Coklat Ice	19000	267
Caramel Coklat Ice	19000	99
V60 (Guess Beans)	18000	100
Rice Bowl 1 Paper	15000	367
Mie Rebus Topping	15000	96
Mie Goreng Topping	15000	98
Kentang Goreng	15000	367
Sosis Goreng	13000	49
Cireng	13000	65
Otak - Otak	13000	140
Roti Bakar Coklat Keju	15000	100
Jamur Crispy	12000	277
Dimsum	15000	234
Kentang Sepatu Wedges	15000	96

Onion Ring	15000	61
Ricebowl Barbeque	15000	362
Chicken Holy Wings	15000	154
Nasi	4000	106
Topping Telur	4000	82

Langkah awal dalam algoritma k-means adalah menentukan jumlah K atau *cluster*. Kemudian menentukan centroid atau pusat *cluster* secara acak (random). Parameter dalam mengukur kinerja atau akurasi algoritma K-means dilakukan dengan menghitung *Davies Bouldin Index (DBI)*. DBI adalah algoritma yang memiliki hasil *cluster* berdasarkan jarak *inter-cluster* rendah (kemiripan *inter-cluster* tinggi) dan jarak antar *cluster* tinggi (kesamaan *inter-cluster* rendah) akan mempunyai nilai DBI rendah. Kumpulan *cluster* dengan nilai DBI terendah akan dianggap sebagai algoritma terbaik. Dengan melakukan evaluasi berdasarkan jumlah K, yaitu K = 2, K = 3 dan K = 4, maka nilai DBI yang didapat sebagai berikut:

- K = 2 : -0,539
- K = 3 : -0,457
- K = 4 : -0,471

Berdasarkan hasil tersebut, maka *cluster* yang dipilih *cluster* berjumlah 3 (K = 3), yang terdiri dari *cluster* 1 (C1) merupakan menu dengan tingkat penjualan rendah, *cluster* 2 (C2) merupakan menu dengan tingkat penjualan sedang, dan *cluster* 3 (C3) merupakan menu dengan tingkat penjualan tinggi.

Penentuan awal *cluster* ini dapat diasumsikan sebagai berikut:

Centroid 1, diambil dari data ke- 1 : (216, 15000)

Centroid 2, diambil dari data ke- 27 : (246, 15000)

Centroid 3, diambil dari data ke- 48 : (277, 12000)

Langkah selanjutnya adalah menghitung jarak antara data dengan centroid awal menggunakan persamaan Euclidean distance. Kemudian mengelompokkan data ke dalam *cluster* berdasarkan data yang memiliki jarak terdekat maka data tersebut akan masuk kedalam *cluster* tersebut yang dijelaskan pada Tabel 2.

TABEL 2.
HASIL JARAK TERDEKAT DAN CLUSTER (ITERASI 1)

Product	C1	C2	C3	Jarak Terdekat	Class
V60	0	30	3001	0	1
Tubruk	5004	5005	2016	2016	3
Vietnam Drip Hot	2000	2000	1000	1000	3
Vietnam Drip Ice	116	146	3005	116	1
Espresso	5000	5000	2002	2002	3
Americano Hot	2001	2003	1009	1009	3
Americano Ice	159	189	3008	159	1
Cappucino Hot	126	156	3006	126	1
Cappucino Ice	1015	1010	4002	1010	2
Cafe Latte Hot	116	86	3001	86	2

Cafe Latte Ice	1012	1008	4001	1008	2
Caramel latte Hot	156	186	3008	156	1
Caramel latte Ice	1008	1005	4001	1005	2
Hazelnut Latte Hot	162	192	3008	162	1
Hazelnut Latte Ice	1005	1003	4000	1003	2
Vanilla Latte Hot	204	234	3012	204	1
Vanilla Latte ice	1031	1024	4005	1024	2
KOPASUS	558	528	3041	528	2
Air Mineral 350ml	10032	10030	7039	7039	3
Moccacino Hot	1015	1021	2014	1015	1
Moccacino Ice	163	193	3008	163	1
Matcha Hot	1021	1027	2018	1021	1
Matcha Ice	553	523	3040	523	2
Coklat Hot	1011	1016	2011	1011	1
Coklat Ice	218	188	3004	188	2
Butter Bir	201	231	3011	201	1
Kopi Bir	30	0	3000	0	2
Lemon tea	5056	5051	2115	2115	3
Geraldine	5001	5002	8002	5001	1
Japanese Ice	126	156	3006	126	1
Caramel Machiato Ice	8002	8002	11002	8002	1
Jasmine Latte	3003	3005	6003	3003	1
Rajampat	5000	5000	8000	5000	2
Blue Bottels	5003	5004	8004	5003	1
Susu Putih	5001	5001	2005	2005	3
Lychee Tea	2000	2000	5000	2000	2
Hazelnut Coklat Ice	4000	4000	7000	4000	2
Caramel Coklat Ice	4002	4003	7002	4002	1
V60 (Guess Beans)	3002	3004	6003	3002	1
Rice Bowl 1 Paper	151	121	3001	121	2
Mie Rebus Topping	120	150	3005	120	1
Mie Goreng Topping	118	148	3005	118	1
Kentang Goreng	151	121	3001	121	2
Sosis Goreng	2007	2010	1026	1026	3
Cireng	2006	2008	1022	1022	3
Otak - Otak	2001	2003	1009	1009	3
Roti Bakar Coklat Keju	116	146	3005	116	1
Jamur Crispy	3001	3000	0	0	3
Dimsum	18	12	3000	12	2
Kentang Sepatu Wedges	120	150	3005	120	1
Onion Ring	155	185	3008	155	1
Ricebowl Barbeque	146	116	3001	116	2
Chicken Holy Wings	62	92	3003	62	1
Nasi	11001	11001	8002	8002	3
Topping Telur	11001	11001	8002	8002	3

Setelah itu menentukan centroid baru dengan menghitung rata-rata data yang ada dimasing – masing *cluster*. Lalu ulangi kembali menghitung jarak terdekat dengan centroid baru untuk tahapan iterasi selanjutnya.

Proses iterasi akan terus dilakukan hingga nilai centroid serta letak atau posisi *cluster* tidak berubah atau berpindah lagi. Pada penelitian ini iterasi dilakukan hingga iterasi 5. Berikut adalah hasil perhitungan pengulangan atau iterasi ke-5, dapat dilihat pada Tabel 3 berikut ini :

TABEL 3.
HASIL ITERASI KE 5

Product	C1	C2	C3	Jarak Terdekat	Clas s
V60	4626	225	7430	225	2
Tubruk	9626	4779	2451	2451	3
Vietnam Drip Hot	6626	1776	5430	1776	2
Vietnam Drip Ice	4625	251	7433	251	2
Espresso	9625	4775	2435	2435	3
Americano Hot	6625	1776	5433	1776	2
Americano Ice	4626	272	7435	272	2
Cappucino Hot	4625	255	7433	255	2
Cappucino Ice	3635	1238	8429	1238	2
Cafe Latte Hot	4630	256	7429	256	2
Cafe Latte Ice	3633	1235	8429	1235	2
Caramel latte Hot	4625	271	7435	271	2
Caramel latte Ice	3632	1232	8429	1232	2
Hazelnut Latte Hot	4626	274	7435	274	2
Hazelnut Latte Ice	3630	1230	8429	1230	2
Vanilla Latte Hot	4626	300	7437	300	2
Vanilla Latte ice	3641	1252	8429	1252	2
KOPASUS	4670	607	7440	607	2
Air Mineral 350ml	14652	9808	2655	2655	3
Moccacino Hot	5626	793	6436	793	2
Moccacino Ice	4626	275	7435	275	2
Matcha Hot	5626	800	6438	800	2
Matcha Ice	4670	602	7440	602	2
Coklat Hot	5625	788	6435	788	2
Coklat Ice	4635	317	7429	317	2
Butter Bir	4626	298	7437	298	2
Kopi Bir	4627	228	7429	228	2
Lemon tea	9661	4834	2503	2503	3
Geraldine	376	5226	12431	376	1

Japanese Ice	4625	255	7433	255	2
Caramel Machiato Ice	3376	8227	15432	3376	1
Jasmine Latte	1626	3228	10432	1626	1
Rajampat	403	5225	12429	403	1
Blue Bottels	385	5228	12433	385	1
Susu Putih	9625	4776	2439	2439	3
Lychee Tea	2628	2225	9429	2225	2
Hazelnut Coklat Ice	641	4225	11429	641	1
Caramel Coklat Ice	626	4226	11432	626	1
V60 (Guess Beans)	1625	3227	10432	1625	1
Rice Bowl 1 Paper	4631	274	7429	274	2
Mie Rebus Topping	4625	252	7433	252	2
Mie Goreng Topping	4625	251	7433	251	2
Kentang Goreng	4631	274	7429	274	2
Sosis Goreng	6625	1782	5437	1782	2
Cireng	6625	1781	5437	1781	2
Otak - Otak	6625	1776	5433	1776	2
Roti Bakar Coklat Keju	4625	251	7433	251	2
Jamur Crispy	7627	2776	4429	2776	2
Dimsum	4626	226	7430	226	2
Kentang Sepatu Wedges	4625	252	7433	252	2
Onion Ring	4625	270	7435	270	2
Ricebowl Barbeque	4631	271	7429	271	2
Chicken Holy Wings	4625	232	7431	232	2
Nasi	15625	10776	3580	3580	3
Topping Telur	15625	10776	3582	3582	3

Pada iterasi kelima, hasil data yang dikelompokkan sama dengan hasil pengelompokkan iterasi sebelumnya atau keempat, dengan demikian proses pengolahan dataset dihentikan. Hasil dari pengolahan dataset didapatkan *cluster* 1 sebanyak 8 menu, *cluster* 2 sebanyak 40 menu dan *cluster* 3 sebanyak 7 menu.

Pada penelitian ini, akan mengimplementasi algoritma k-means menggunakan *tools* atau alat *RapidMiner* dengan menggunakan nilai $K = 2$ sampai dengan $K=4$ yang kemudian dilakukan evaluasi dengan uji validasi *clustering* yaitu *Daives Bouldin Index (DBI)* untuk mengidentifikasi atau mengetahui *cluster* terbaik.

Cluster Model

Cluster 0: 40 items

Cluster 1: 7 items

Cluster 2: 8 items

Total number of items: 55

Gambar 2. Hasil Cluster Algoritma K-means

Gambar diatas menunjukkan hasil pengolahan data penjualan di Dpom Coffee berdasarkan Algoritma K-means yang dibagi menjadi 3 *cluster* dengan nilai uji validasi terendah berdasarkan DBI yaitu sebesar -0,457, menghasilkan *Cluster* 0 sebanyak 40 menu, *Cluster* 1 sebanyak 7 menu dan *cluster* 2 sebanyak 8 menu. Dimana berdasarkan centroid akhir menunjukkan bahwa *Cluster* 0 merupakan *cluster* sedang, *Cluster* 1 merupakan *cluster* tinggi dan *Cluster* 2 merupakan *cluster* rendah.

TABEL 4.
HASIL CENTROID AKHIR

Atribut	Cluster 0	Cluster 1	Cluster 2
Total	210.275	359.143	125.5
Harga	14775	7571.429	19625

V. KESIMPULAN

Berdasarkan hasil pengolahan dataset penjualan di Dpom Coffee menggunakan *tools Rapidminer*, menghasilkan 3 *cluster* dengan nilai DBI sebesar -0.457, yaitu dengan cluster 1 menu dengan tingkat penjualan rendah sebanyak 8 yaitu, Geraldine, Caramel Machiato Ice, Jasmine Latte, Raja Ampat, Blue Bottles, Hazelnut Coklat Ice, Caramel Coklat Ice dan V60(Guesss beans). Cluster 2 dengan menu yang memiliki tingkat penjualan sedang sebanyak 40 yaitu, V60, Vietnam Drip(Hot/Ice), Americano(Hot/Ice), Cappuchino (Hot/Ice), CoffeeshopLatte (Hot/Ice), Vanilla Late (Hot/Ice), Kopasus, Moccachino (Hot/Ice), Matcha (Hot/Ice), Coklat (Hot/Ice), Butter Bir, Kopi Bir, Japanese Ice, Lychee Tea, Rice Bowl Black Paper, Mie Rebus Topping, Mie Goreng Topping, Kentang Goreng, Sosis Goreng, Cireng, Otak –otak, Roti bakar Coklat Keju, Jamur Crispy, Dimsum, Kentang Sepatu Wedges, Onion Ring, RiceBowl Barbequw, dan Chicken holywings. Sedangkan untuk Cluster 3 sebanyak 7 Menu dengan tingkat penjualan tinggi yaitu, Air Mineral, Espresso, Lemon tea, Susu putih, Tubruk, Nasi, dan Topiing Telur.

Dalam hasil tersebut masih ada beberapa menu yang tingkat penjualannya masih rendah. Dengan demikian hal tersebut dapat dijasikan sebagai acuan bagi Dpom Coffee untuk mengevaluasi, mengatur strategi penjualan atau mengatur stok bahan menu di Dpom Coffee.

Dengan demikian, penerapan data mining menggunakan algoritma k-means clustering dapat dimanfaatkan dalam menjalankan bisnis *coffeeshop* untuk mengetahui pengelompokan menu makanan dan minuman mana yang diminati atau kurang diminati berdasarkan harga dan jumlah penjualan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada pihak Dpom Coffee yang telah memberikan izin untuk akses data penjualan dan menjadikan Dpom Coffee sebagai tempat objek penelitian.

DAFTAR PUSTAKA

- [1] D. Kharismal and N. Nurkomalasari, "Penerapan Strategi Bauran Pemasaran Di Umkm Katering Di Semarang Barat," *Gemawisata: Jurnal Ilmiah Pariwisata*, vol. 17, no. 2, Art. no. 2, May 2021, Accessed: May 09, 2022. [Online]. Available: <http://stiepari.greenfrog-ts.co.id/jurnal/index.php/JT/article/view/250>
- [2] S. Nurajizah, "Analisa Transaksi Penjualan Obat menggunakan Algoritma Apriori | Nurajizah | INOVTEK Polbeng - Seri Informatika," vol. 4, no. 1, pp. 35–44, 2019, doi: <https://doi.org/10.35314/isi.v4i1.938>.
- [3] Normah, I. Yulianti, D. Novianti, M. N. Winnarto, A. Zumarniansyah, and S. Linawati, "Comparison of Classification C4.5 Algorithms and Naïve Bayes Classifier in Determining Merchant Acceptance on Sponsorship Program," *J. Phys.: Conf. Ser.*, vol. 1641, no. 1, p. 012006, Nov. 2020, doi: 10.1088/1742-6596/1641/1/012006.
- [4] D. E. Kurniawan and A. Fatulloh, "Clustering of Social Conditions in Batam, Indonesia Using K-Means Algorithm and Geographic Information System," *International Journal of Earth Sciences and Engineering (IJEE)*, vol. 10, no. 5, pp. 1076–1080, 2017.
- [5] I. S. M. Negara, P. Purwono, and I. A. Ashari, "Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 dengan Algoritma K-means," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 6, no. 3, Art. no. 3, Sep. 2021, doi: 10.31328/jointecs.v6i3.2693.
- [6] R. Kesuma Dinata, H. Novriando, N. Hasdyna, and S. Retno, "Reduksi Atribut Menggunakan Information Gain untuk Optimasi Cluster Algoritma K-Means | Dinata | JEPIN (Jurnal Edukasi dan Penelitian Informatika)," <https://jurnal.untan.ac.id/index.php/jepin/article/view/37606> (accessed May 09, 2022).
- [7] A. S. Mardani, "Strategi Marketing dalam Pengembangan UMKM Kopi Jetak Mentari Menurut Perspektif Syariah (Studi Kasus pada UMKM Kopi Jetak Mentari di Desa Kedungdowo)," skripsi, IAIN KUDUS, 2020. Accessed: May 09, 2022. [Online]. Available: <http://repository.iainkudus.ac.id/4417/>
- [8] H. Halma, "Analisis Penjualan Minuman Thai Tea Pada Toko IL.SHOP18 Di Sangatta.," *Madani Accounting And Management Journal*, vol. 7, no. 2, Art. no. 2, Sep. 2021, doi: 10.51882/jamm.v7i2.40.
- [9] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.
- [10] A. A. Fajrin and N. Mirantika, "Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan," *Buffer Informatika*, vol. 6, no. 1, pp. 1–14, 2020.
- [11] A. A. Fajrin and A. Maulana, "Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor," *Kumpulan Jurnal Ilmu Komputer (KLIK)*, vol. 5, no. 01, pp. 1–10, 2018.
- [12] P. Prasetyawan, I. Ahmad, R. I. Borman, Y. A. Pahlevi, and D. E. Kurniawan, "Classification of the Period Undergraduate Study Using Back-propagation Neural Network," 2018, pp. 1–5.
- [13] E. Fammaldo and L. Hakim, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Kesejahteraan Keluarga Untuk Program Kartu Indonesia Pintar," *Jurnal Ilmiah Teknologi Infomasi Terapan*, vol. 5, no. 1, pp. 23–31, 2018.
- [14] F. A. Bramasta and R. Halilintar, "Penerapan Data Mining Untuk Menentukan Strategi Penjualan Toko Sepatu," 2021, vol. 5, no. 2, pp. 236–241.
- [15] M. N. V. Waworuntu and M. F. Amin, "Penerapan Metode K-Means Untuk Pemetaan Calon Penerima Jamkesda," *KLIK-Kumpulan Jurnal Ilmu Komputer*, vol. 5, no. 2, pp. 190–200, 2018.
- [16] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient di Puskesmas Karangasambung," *Journal of Applied Informatics and Computing*, vol. 5, no. 2, Art. no. 2, Oct. 2021, doi: 10.30871/jaic.v5i1.3237.
- [17] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *Jurnal Informatika*, vol. 4, no. 01, pp. 20–24, 2019.
- [18] F. Mahmuda, M. A. R. Sitorus, H. Widyastuti, and D. E. Kurniawan, "Clustering Profil Pengunjung Perpustakaan Menggunakan Algoritma K-Means," *Journal of Applied Informatics and Computing*, vol. 1, no. 1, Art. no. 1, 2017, doi: 10.30871/jaic.v1i1.476.
- [19] I. F. Ashari, R. Banjarnahor, D. R. Farida, S. P. Aisyah, A. P. Dewi, and N. Humaya, "Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies," *Journal of Applied Informatics and Computing*, vol. 6, no. 1, Art. no. 1, Jul. 2022, doi: 10.30871/jaic.v6i1.3485.
- [20] Y. P. Sari, A. Primajaya, and A. S. Y. Irawan, "Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang," *INOVTEK Polbeng-Seri Informatika*, vol. 5, no. 2, pp. 229–239, 2020.