

PENERAPAN DATA MINING UNTUK CLUSTERING DENGAN METODE K-MEANS DALAM MENENTUKAN STOK BARANG DI ERA COVID 19 (Studi Kasus Supermarket Kem Chicks)

Pandi Zulfikar^{1*}, Aden², Andi Nur Rahman³
Program Studi Matematika, FMIPA, Universitas Pamulang^{1,2,3}
*Pandizulfikarunpam2017@gmail.com

Penelitian ini bertujuan untuk mengetahui pengelompokan produk paling, sedang, sedikit peminatnya untuk menentukan jumlah stok barang. Selain itu tujuan berikutnya adalah mengetahui karakteristik setiap cluster dan mengetahui variabel apa yang mempengaruhi terbentuknya *cluster*. Sedangkan untuk metode penelitian yang digunakan adalah k-means cluster. Setelah penelitian dilakukan terbentuklah *cluster* pada penelitian ini, dengan uraian *cluster* 1 yang paling diminati dengan jumlah produk hanya 1 yang menandakan stok barang pada *cluster* ini harus lebih banyak dari *cluster* lainnya, *cluster* 2 yang sedang peminatnya dengan jumlah produk 6 yang menandakan bahwa stok barang pada *cluster* ini sedang, *cluster* 3 kurang diminati dengan jumlah produk 119 yang menandakan bahwa stok barang pada *cluster* ini lebih sedikit dari *cluster* lainnya. Karakteristik setiap *cluster* setelah dilakukan penelitian ini, karakteristik yang terdapat dalam *cluster* 1 berisi variabel jumlah barang masuk di bawah rata-rata, jumlah barang yang keluar di bawah rata-rata, dan jumlah sisa stok barang di atas rata-rata, karakteristik yang terdapat pada *cluster* 2 berisi variabel jumlah stok barang yang masuk, barang yang keluar dan sisa barang berada di atas rata-rata, karakteristik yang terdapat pada *cluster* 3 berisi variabel jumlah stok barang yang masuk, barang yang keluar dan sisa barang berada di bawah rata-rata. Sedangkan variabel yang memberikan pengaruh besar dalam terbentuknya *cluster* adalah variabel barang masuk (X_1) dengan nilai F sebesar 143,323 dan nilai signifikan 0,000.

Kata Kunci – Clustering, data mining, k-means, Knowledge Discovery Database, Penyediaan Barang

I. PENDAHULUAN

Pada bulan Maret tahun 2020, Indonesia digemparkan dengan timbulnya suatu virus corona yang berasal dari Cina. Virus corona ini yang dimulai terletak di Wuhan, setelah itu menyebar ke bermacam negeri, termasuk Indonesia. Pemerintah Republik Indonesia lewat Presiden Joko Widodo menginformasikan telah ditemui permasalahan 2 orang yang teridentifikasi virus COVID-19 pada Maret 2020 di Kota Depok. Permasalahan positif di Indonesia kemudian mulai ditemukan di bermacam kota yang lain. Bersamaan dengan penetapan status pandemi oleh World Health Organization pada Maret 2020, hingga bermacam negeri di dunia mulai mempraktikkan *lockdown* buat memutus rantai penyebaran virus. Mengingat kalau keadaan Negeri Indonesia yang ialah negeri kepulauan dengan jumlah penduduk yang besar serta ciri wilayah yang bermacam-macam, hingga pemerintah tidak memilah agar turut mempraktikkan *lockdown*.

Dengan dibatasinya bermacam aktivitas di tempat umum sangat memberikan pengaruh besar bagi sektor perdagangan dan perekonomian. Pandemi COVID-19 bawa perubahan serta akibat yang besar untuk UMKM, usaha restoran, ataupun toko ritel berskala besar sehingga tidak sedikit usaha yang hadapi penyusutan penjualan apalagi wajib menutup usahanya, tetapi terdapat sebagian retail lain semacam supermarket

ataupun mini market yang masih senantiasa dapat beroperasi.

Para petinggi perusahaan berhadapan sebuah sistem yang dapat menghasilkan sebuah informasi agar dapat digunakan dan membantu para petinggi perusahaan untuk menentukan stok barang. Para Eksekutif perusahaan mau mengetahui produk apa saja yang mesti di tingkatkan stoknya, seberapa besar pencapaian hasil yang diperoleh oleh perusahaan. Agar terpenuhinya setiap kebutuhan para konsumen, ada berbagai cara yang dapat dilakukan. Salah satunya adalah dengan melakukan pemanfaatan data perusahaan (*Data Mining*).

Data mining adalah pengambilan data yang tertanam pada database. *Clustering* adalah sebuah tahap pengelompokan yang berlandaskan atas prinsip kesetaraan kelas dan mengurangi kesetaraan antar kelas. Berbagai algoritma pada *clustering* telah berkembang dalam mencapai kinerja yang baik. Keakuratan perkiraan penjualan memiliki dampak yang besar pada penjualan.

Penentuan stok barang yang tidak tepat dikarenakan harus berdasarkan informasi dari jumlah data transaksi penjualan yang besar. dengan hal ini agar memperoleh pengetahuan maka dibutuhkannya beberapa proses dalam mengolah data historis transaksi besar diperlukannya suatu teknik *data mining*. Metode *data mining* yang akan digunakan dalam riset ini yaitu *clustering* dengan metode *K-Means*. Riset ini diharapkan bisa menghasilkan mengcluster atau pengelompokan barang yang harus mempunyai stok banyak sebab sangat banyak peminatnya, stok sedang produk diminati dan stok sedikit sebab produk tidak banyak diminati dari beberapa produk yang banyak yang memudahkan Kem Chicks mengimplementasikan hasil yang di peroleh serta mengetahui jumlah stok barang.

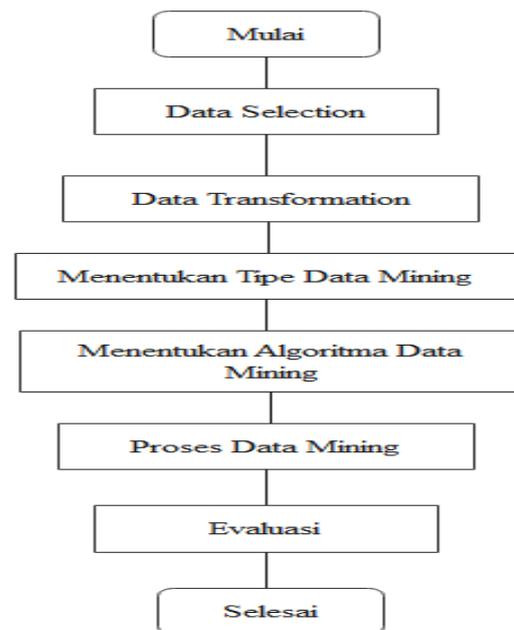
Selain itu penelitian ini juga diharapkan bisa memberikan informasi berupa karakteristik dari *cluster* yang terbentuk dan juga kita dapat mengetahui variabel barang apakah yang berpengaruh besar terhadap

terbentuknya *cluster* dalam pengelompokan barang.

Menurut (Sugiyono 2016) variabel penelitian merupakan sebuah atribut ataupun karakter ataupun *value* pada orang, haluan ataupun aktivitas yang memiliki tipe tertentu yang ditentukan oleh peneliti agar diamati dan kemudian di tarik kesimpulannya. Menurut (Sugiyono 2016) dalam (Kerlinger, 1973) variabel merupakan konstruk (*constructs*) atau yang akan dipelajari. Berikut adalah variabel yang digunakan dalam penelitian ini yaitu:

- Y = Item Barang (126 item)
- X₁ = Jumlah Barang Masuk
- X₂ = Jumlah Barang Terjual
- X₃ = Jumlah Sisa Barang

Berikut adalah alur penelitian ini :



Gambar 1: Alur Penelitian

II. HASIL DAN PEMBAHASAN

Data yang digunakan adalah data penjualan Supermarket Kem Chicks Pacific Place. Pada proses pengambilan data ada beberapa tahapan yang di lakukan sebelum memulai penelitian tersebut. Tahap *Selection* dan *Addition*. Tahap *Selection* (seleksi/pemilihan) data dari merupakan sekumpulan data operasional, dimana peneliti hanya

mengambil data historis diambil dari data transaksi penjualan Supermarket Kem Chicks pada periode 1 Januari 2020 – 31 Desember 2020. Penelitian ini memfokuskan pada produk departemen meat. Tahap *Prepossessing* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

Tahap kedua adalah tahap *Transformation*. Pada tahap proses *transformasi* data peneliti dengan cara mengubah nama produk menjadi kode produk agar mempermudah dalam memproses data sehingga didapatkannya kode baru untuk produk. Pada tahap ini juga dilakukan proses standarisasi data. Berikut hasil dari transformasi data.

Tabel 1: Tahap Transformasi Data

Variabel	X ₁	X ₂	X ₃
Y ₁	-35367	-35838	.31334
Y ₂	-39941	-39960	-.11343
Y ₃	-.51440	-.51492	-.12210
Y ₄	-.45535	-.45635	-.05852
Y ₅	-.51069	-.50617	-.58019
Y ₆	-.25902	-.25211	-.71586
Y ₇	-.27901	-.36075	736.265
Y ₈	-.47384	-.47182	-.34047
Y ₉	-.27815	-.27452	-.42251
Y ₁₀	-.21362	-.20701	-.67315
Y ₁₁	-.27502	-.26992	-.55594
Y ₁₂	-.56826	-.56389	-.58516
Y ₁₃	.43386	.43410	.12019
Y ₁₄	-.02658	-.02522	-.13270
Y ₁₅	.01286	.01426	-.12339
Y ₁₆	-.52572	-.52540	-.20206
Y ₁₇	-.47940	-.47769	-.31381
Y ₁₈	-.50743	-.51439	.46796
Y ₁₉	.24878	.25498	-.48417
Y ₂₀	194.365	195.603	-.48963
Y ₂₁	-.37016	-.36521	-.57360
Y ₂₂	-.56455	-.55872	-.71763

Y ₂₃	-.48690	-.48282	-.53202
Y ₂₄	-.55724	-.55170	-.68856
Y ₂₅	.34183	.34357	-.04600
Y ₂₆	.12114	.11989	.15374
Y ₂₇	528.945	530.579	.24912
Y ₂₈	161.062	160.841	.73080
Y ₂₉	305.794	306.624	.24912
Y ₃₀	-.55901	-.55367	-.67074
Y ₃₁	338.688	334.090	530.682
Y ₃₂	-.54633	-.53607	-111.566
Y ₃₃	.38123	.37888	.33983
Y ₃₄	-.29692	-.30087	.26293
Y ₃₅	-.01620	-.01862	.21588
Y ₃₆	170.136	170.317	.39443
Y ₃₇	-.56093	-.55724	-.52094
Y ₃₈	-.27043	-.27616	.43408
Y ₃₉	.86208	.86110	.37307
Y ₄₀	-.55629	-.55149	-.62033
Y ₄₁	.51307	.49235	205.833
Y ₄₂	.08874	.07664	113.221
Y ₄₃	.68082	.66822	137.305
Y ₄₄	-.38105	-.38460	.19886
Y ₄₅	.00736	.00641	.08904
Y ₄₆	-.30930	-.30499	-.49493
Y ₄₇	-.48999	-.48454	-.65854
Y ₄₈	-.40455	-.40262	-.30916
Y ₄₉	-.53054	-.52550	-.63397
Y ₅₀	-.42474	-.42405	-.20270
Y ₅₁	-.31622	-.31106	-.57457
Y ₅₂	-.51537	-.51488	-.21426
Y ₅₃	127.905	128.093	.24912
Y ₅₄	-.46675	-.46237	-.55225
Y ₅₅	-.48786	-.47991	-.88541
Y ₅₆	-.12009	-.11678	-.34095
Y ₅₇	-.31850	-.31144	-.74942
Y ₅₈	-.48368	-.47670	-.79614
Y ₅₉	-.49676	-.49633	-.20319
Y ₆₀	-.42306	-.42324	-.12258
Y ₆₁	-.48104	-.47816	-.42155

Y ₆₂	-.48350	-.48285	-.21860
Y ₆₃	-.40360	-.40514	.00763
Y ₆₄	-.42815	-.42992	.02064
Y ₆₅	-.40511	-.41986	121.249
Y ₆₆	-.47631	-.47446	-.32521
Y ₆₇	-.39308	-.39178	-.24750
Y ₆₈	-.31411	-.31399	-.11407
Y ₆₉	-.53054	-.53607	.32940
Y ₇₀	-.49165	-.49614	.24751
Y ₇₁	554.783	555.639	104.358
Y ₇₂	.25841	.25397	.48980
Y ₇₃	-.46017	-.46343	.14620
Y ₇₄	194.020	193.696	.93295
Y ₇₅	-.03114	-.03168	.03942
Y ₇₆	313.315	314.006	.40037
Y ₇₇	.27507	.28309	-.64088
Y ₇₈	.36271	.36491	-.08132
Y ₇₉	-.49228	-.48456	-.86615
Y ₈₀	.17708	.17841	-.06302
Y ₈₁	.32772	.32904	-.01212
Y ₈₂	-.46796	-.47387	.38431
Y ₈₃	.21482	.22145	-.53410
Y ₈₄	.04443	.04920	-.41994
Y ₈₅	.80361	.80386	.24109
Y ₈₆	-.57054	-.56254	-.91720
Y ₈₇	-.48642	-.48836	.01694
Y ₈₈	-.21773	-.23229	125.648
Y ₈₉	-.44318	-.44843	.33325
Y ₉₀	-.19333	-.18826	-.52624
Y ₉₁	.11024	.10470	.54198
Y ₉₂	-.54649	-.54262	-.53282
Y ₉₃	.27980	.27313	.70030
Y ₉₄	.84041	.84889	-.49701
Y ₉₅	.06456	.06096	.34931
Y ₉₆	.06910	.07761	-.75391
Y ₉₇	-.55686	-.54663	-111.566
Y ₉₈	-.56738	-.55719	-111.566
Y ₉₉	-.45160	-.44099	-111.566
Y ₁₀₀	.10859	.11307	-.37274

Y ₁₀₁	-.18255	-.18629	.28123
Y ₁₀₂	-.45982	-.46045	-.09416
Y ₁₀₃	-.55008	-.54775	-.39377
Y ₁₀₄	-.56500	-.55988	-.65260
Y ₁₀₅	-.55525	-.55149	-.52559
Y ₁₀₆	-.04264	-.04924	.58806
Y ₁₀₇	-.15431	-.15494	.00699
Y ₁₀₈	.32908	.32082	.86182
Y ₁₀₉	-.27202	-.27839	.49124
Y ₁₁₀	-.53817	-.53237	-.70623
Y ₁₁₁	-.25193	-.24839	-.40549
Y ₁₁₂	-.29579	-.29774	.08133
Y ₁₁₃	-.41219	-.40624	-.67813
Y ₁₁₄	-.35305	-.36164	.66706
Y ₁₁₅	-.46796	-.45927	-.94707
Y ₁₁₆	-.56262	-.55542	-.84158
Y ₁₁₇	-.48717	-.48359	-.48674
Y ₁₁₈	-.56448	-.55730	-.84030
Y ₁₁₉	-.35702	-.35534	-.27078
Y ₁₂₀	-.52017	-.52601	.36119
Y ₁₂₁	-.45052	-.45396	.16611
Y ₁₂₂	-.44350	-.44763	.23033
Y ₁₂₃	-.46140	-.46720	.37692
Y ₁₂₄	-.46184	-.46324	-.02448
Y ₁₂₅	-.06799	-.07066	.22086
Y ₁₂₆	-.40491	-.41507	.79358

Tabel 2. Descriptive Statsitics

	N	Mini mum	Maxi mum	Mea n	Std. Deviasi on
X1	126	8.40	6983.56	658.8367	1140.03635
X2	126	4.39	6956.66	644.9398	1135.93961
X3	126	.00	105.61	13.8969	12.45625
Valid N (listwise)	126				

Berdasarkan hasil analisa berikut diperoleh, ada beberapa yang dianalisa yaitu jumlah fakta, minimal, maksimal, rata-rata, dan simpangan baku. Pada variabel X_1 terdapat data berjumlah 126 dengan nilai minimal 8,40, maksimal 6983,56, rata-rata 658,8367 dan simpangan baku 1140,03635. Pada variabel X_2 terdapat data berjumlah 126 dengan nilai minimal 4,39, maksimal 6956,66, rata-rata 644,9398 dan simpangan baku 1135,93961. Pada variabel X_3 terdapat data berjumlah 126 dengan nilai minimal 0, maksimal 105,61, rata-rata 13, 8969 dan simpangan baku 12,45625.

Tahap ke tiga menentukan jenis *data mining* yang tepat. Jenis *Data Mining* yang digunakan pada riset ini yaitu *clustering* untuk mengelompokan barang dalam menentukan stok barang pada kriteria stok barang kurang, sedang dan paling banyak minatnya.

Tahap ke empat *Data Mining* menentukan algoritma *data mining*. Metode yang digunakan pada riset ini adalah metode *K-Means* yang mengkluster barang menjadi 3 kelompok untuk stok banyak berdasarkan banyaknya permintaan pelanggan terhadap barang tersebut.

Tahap ke lima tahap *Data Mining* proses menggunakan algoritma *data mining*. Dalam proses ini dilaksanakannya proses pengaplikasian proses algoritma *data mining* yang sudah dipilih sebelumnya.

Tabel 3: Tabel Initial Cluster

	Cluster		
	1	2	3
Zscore(X1)	-.27901	5.54783	-.56738
Zscore(X2)	-.36075	5.55639	-.55719
Zscore(X3)	7.36265	1.04358	-1.11566

Menurut (Santoso 2017) hasil dari *output Initial Cluster Centers* Adalah tampilan pertama tahap *clustering* fakta ke-1 sebelum data berikut dilaksanakan pengulangan, pada data berikut merupakan tahap dalam pembentukan 3 cluster. Namun, tabel *Initial Cluster Centers* tidak dianalisa.

Tabel 4: Tabel Iteration History

Iteration	Change in Cluster Centers		
	1	2	3
1	.000	2.118	1.140
2	.000	.513	.027
3	.000	.000	.000

Menurut (Santoso 2017) Tampilan *Iteration Cluster* adalah tahap iterasi pada pembagian *cluster* dari tabel initial dan proses iterasi yang dihasilkan sebanyak 3 kali. Proses iterasi pertama dan kedua terdapat centeroid yang tidak signifikan dan proses pengulangan ketiga terbentuk centeroid signifikan. Jadi, setelah seluruh *cluster* telah terbentuk lalu pengulangan terhenti di pengulangan ke-3 dan jarak minimalnya 8,485.

Tabel 5: Tabel *Final Cluster Centers*

Final Cluster Centers			
	Cluster		
	1	2	3
Zscore(X1)	-.27901	3.72591	-.18552
Zscore(X2)	-.36075	3.72439	-.18475
Zscore(X3)	7.36265	1.36366	-.13063

Tabel *Final Cluster Centers* merupakan hasil akhir dari proses pada *clustering* yang membuat *cluster* sebanyak 3 pada setiap variabel. Variabel dalam tabel pusat *cluster* terakhir adalah hasil nilai standarisasi. Nilai-nilai yang tercipta pada pusat *cluster* terakhir memiliki makna. Menurut (Santoso 2017) nilai minus mempunyai makna jika data berada di bawah rata-rata keseluruhan lalu nilai plus mempunyai makna jika data berada di atas rata-rata keseluruhan.

Berdasarkan hasil pengolahan data diatas dari setiap variabel, didapatkan spesifikasi dalam setiap *cluster*. Berikut merupakan penjabaran pada setiap *cluster*:

1. *Cluster 1*

Spesifikasi dalam *cluster* ke-1 terdapat variabel jumlah barang masuk di bawah rata-rata, jumlah barang yang keluar di bawah rata-rata, dan jumlah sisa stok barang di atas rata-rata.

2. *Cluster 2*

Spesifikasi dalam *cluster* ke-2 terdapat variabel dengan besaran stok barang yang masuk, barang yang keluar dan sisa barang berada di atas rata-rata.

3. *Cluster 3*

Spesifikasi dalam *cluster 3* terdapat variabel dengan besaran stok barang yang masuk di, barang yang keluar dan sisa barang berada di bawah rata-rata.

Tabel 6: Tabel Anova

	Cluster		Error		F	Sig.
	Mean Square	Df	Mean Square	Df		
Zscore(X1)	43.734	2	.305	123	143.323	.000
Zscore(X2)	43.709	2	.306	123	143.056	.000
Zscore(X3)	33.698	2	.468	123	71.956	.000

Pada tabel anova dapat dilihat bahwa f , semakin besar nilai f semakin besar perbedaan pada *cluster 1*, *cluster 2*, dan *cluster 3*. Dan jika nilai signifikansi di bawah 0,05 menandakan bahwa *cluster 1*, *cluster 2*, *cluster 3* berbeda secara signifikan. Dengan nilai F sebesar 143,323 dengan nilai signifikan 0,000 maka variabel (X_1) barang masuk maka variabel ini adalah variabel yang memiliki pengaruh terhadap ketiga *cluster* yang terbentuk.

Tabel 6: Tabel Number of Cases in Each Cluster

Number of Cases in each Cluster	
Cluster	1
	2
	3
Valid	126.000
Missing	.000

Pada Tabel *Number of Cases in each Cluster*, tidak terdapat kesalahan data lalu pada *cluster 1* memiliki anggota berjumlah 1 produk, *cluster 2* sebanyak 6 produk, *cluster 3* sebanyak 119 produk dengan penjabaran sebagai berikut:

1. *Cluster 1*
Yaitu *cluster* yang kurang diminati terdapat 1 produk yang menandakan produk tersebut stoknya lebih sedikit dari *cluster* yang lainnya.
 2. *Cluster 2*
Yaitu *cluster* yang paling banyak peminatnya, terdapat 6 produk yang menandakan produk tersebut stoknya lebih banyak dari *cluster* yang lain.
 3. *Cluster 3*
Yaitu *cluster* yang sedang peminatnya terdapat 119 produk yang menandakan bahwa untuk produk tersebut penyediaan terhadap stok barangnya sama seperti biasanya.
- a. *Cluster 1*
rata-rata, jumlah barang yang keluar di bawah rata-rata, dan jumlah sisa stok barang di atas rata-rata.
 - b. *Cluster 2*
Spesifikasi dalam *cluster* ke-2 terdapat variabel dengan besaran stok barang yang masuk, barang yang keluar dan sisa barang berada di atas rata-rata.
 - c. *Cluster 3*
Spesifikasi dalam *cluster* 3 terdapat variabel dengan besaran stok barang yang masuk, barang yang keluar dan sisa barang berada di bawah rata-rata.
3. Variabel yang membagikan perbandingan sangat besar dalam *cluster* yang tercipta pada permasalahan ini yaitu variabel benda masuk(X1) dengan nilai F sebesar 143,323 serta nilai signifikan 0,000

III. KESIMPULAN

Dari hasil dan pembahasan dalam penelitian yang sudah dilaksanakan, jadi bisa disimpulkan sebagai berikut:

1. Analisis dengan metode K-Means *Cluster* dalam berdasarkan stok barang yang paling, sedang dan kurang maka terbentuklah:
 - a. *Cluster 1*
Yaitu *cluster* 1 yang kurang diminati terdapat 1 produk yang menandakan produk tersebut stoknya lebih sedikit dari *cluster* yang lainnya.
 - b. *Cluster 2*
Yaitu *cluster* yang paling banyak peminatnya, terdapat 6 produk yang menandakan produk tersebut stoknya lebih banyak dari *cluster* yang lain.
 - c. *Cluster 3*
Yaitu *cluster* yang sedang peminatnya terdapat 119 produk yang menandakan bahwa untuk produk tersebut penyediaan terhadap stok barangnya sama seperti biasanya.
2. Karakteristik *cluster* yang terbentuk adalah sebagai berikut:
 - a. *Cluster 1*
Spesifikasi dalam *cluster* ke-1 terdapat variabel jumlah barang masuk di bawah

REFERENSI

- [1] Sugiyono. 2016. *Metode Penelitian Kuantitatif, Kualitatif, Dan R&D*. Bandung: Alfabeta..
- [2] Santoso, Singgih. 2017. *Statistik Multivariat Dengan SPSS - Google Books*. Jakarta: PT Elex Media Komputindo.
https://www.google.co.id/books/edition/Statistik_Multivariat_dengan_SPSS/qi9IDwAAQBAJ?hl=id&gbpv=0. Menggunakan Algoritma Floyd Warshall,” *Jurnal Riset dan Aplikasi Matematika, UNESA Surabaya*, vol. 2, no. 2, pp. 101–111, 2018.