

PEMANFAATAN DATA MINING UNTUK MENGELOMPOKKAN KATEGORI PENJUALAN PRODUK

Herliyani Hasanah¹, Wahyu Larasati²

^{1,2} *Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta*

¹herliyani_hasanah@udb.ac.id

ABSTRAK

Ketersediaan barang dan kelengkapan barang pada suatu toko adalah elemen yang sangat penting. Sehingga proses manajemen untuk mengatur ketersediaan persediaan barang sangat diperlukan untuk menghindari penumpukan barang yang sama dan kurang diminati oleh pelanggan. Selain itu manajemen proses untuk mengatur ketersediaan persediaan barang yang dibutuhkan juga sangat dibutuhkan untuk memaksimalkan barang tertentu yang paling banyak diminati oleh pelanggan. Untuk mengetahui barang yang laris, tidak laris di Restu Baru masih dalam satu *query* di database menggunakan *Navicat*, jadi di *sort by* berdasarkan *query* terbanyak maka dapat disebut barang laris, dan yang paling bawah disebut tidak laris. belum ada pengelompokan tersendiri, selain itu jika menginginkan informasi mengenai barang laris, kurang laris dan tidak laris harus menemui divisi IT.

Data yang digunakan pada penelitian ini diperoleh dengan cara mengambil langsung data hasil penjualan di Restu Baru. Variabel penelitian tersebut meliputi nama produk, tanggal transaksi, jumlah persediaan barang dan jumlah penjualan. Periode waktu pengambilan data adalah bulan Juli 2018. Dalam penelitian ini dilakukan beberapa tahap yang meliputi pengumpulan dan analisis data, perancangan algoritma K-Means Clustering, perancangan proses, simulasi dengan Weka dan pengujian.

Tahap pengujian dilakukan dengan membandingkan hasil perhitungan dengan simulasi Weka. Berdasarkan hasil pengujian berdasarkan perancangan K-Means Clustering dengan simulasi menggunakan Weka menghasilkan keluaran yang sama yaitu cluster 1 terdapat 42%, cluster 2 terdapat 32%, cluster 3 terdapat 26%.

Kata kunci : penjualan, barang, K-Means Clustering

ABSTRACT

The availability of goods and the completeness of goods in a store is a very important element. So that the management process to regulate the availability of inventory is very necessary to avoid the accumulation of goods that are the same and less attractive to customers. In addition, the management process to regulate the availability of the required inventory is also needed to maximize certain items that are most in demand by customers. To find out which items are in demand, they are not in demand in Restu Baru, still in one query in the database using Navicat, so sort by based on the most queries it can be called the bestselling item, and the bottom one is not in demand. information about best-selling items, less in demand and not in demand to meet the IT division.

The data used in this study was obtained by taking direct sales data at Restu Baru. The research variables include product name, transaction date, inventory amount and number of sales. The time period for data collection is July 2018. In this study several stages were conducted which included data collection and analysis, design of the K-Means Clustering algorithm, process design, simulation with Weka and testing.

The testing phase is done by comparing the results of calculations with the Weka simulation. Based on the results of testing based on the design of K-Means Clustering with simulations using Weka to produce the same output, cluster 1 has 42%, cluster 2 has 32%, cluster 3 has 26%.

Keyword: sales, goods, K-Means Clustering

PENDAHULUAN

Pelaku bisnis harus selalu memikirkan cara untuk terus bertahan dan jika mungkin mengembangkan skala bisnis. Untuk mencapai hal tersebut, terdapat tiga kebutuhan bisnis yang dapat dilakukan, yaitu penambahan jenis maupun peningkatan kapasitas produk, pengurangan biaya operasional perusahaan, serta peningkatan efektifitas pemasaran dan keuntungan [1]. Agar bisa memenuhi kebutuhan-kebutuhan bisnis diatas banyak cara yang dapat ditempuh salah satunya adalah dengan melakukan analisis data perusahaan.

Ketersediaan barang dan kelengkapan barang pada suatu toko adalah elemen yang sangat penting. Sehingga proses manajemen untuk mengatur ketersediaan persediaan barang sangat diperlukan untuk menghindari penumpukan barang yang sama dan kurang diminati oleh pelanggan. Selain itu manajemen proses untuk mengatur ketersediaan persediaan barang yang dibutuhkan juga sangat dibutuhkan untuk memaksimalkan barang tertentu yang paling banyak diminati oleh pelanggan.

Restu Baru adalah salah satu contoh perusahaan yang bergerak di bidang ritel khususnya perlengkapan bayi dan anak-anak. Jenis barang pada Restu Baru setiap bulannya selalu bertambah banyak dan type barang tersebut semakin bervariasi, maka semakin banyak pula barang yang dapat ditawarkan pada konsumen. Untuk melakukan analisis data penjualan yaitu mengetahui barang yang laris, tidak laris di Restu Baru masih dalam satu *query* di *database* menggunakan *Navicat*, jadi di *sort by* berdasarkan *query* terbanyak maka dapat disebut barang laris, dan yang paling bawah disebut tidak laris. Saat ini belum ada pengelompokan tersendiri, selain itu jika menginginkan informasi mengenai barang laris, kurang laris dan tidak laris harus menemui divisi IT.

Penelitian yang memanfaatkan clustering di bidang ritel telah dilakukan oleh Giering [2] dan Kusri [3]. Dalam penelitian tersebut, penentuan kelompok produk ke dalam kategori bergerak cepat dan bergerak lambat di industri ritel dilakukan dengan menggunakan pengelompokan proses. Clustering merupakan proses pengelompokan satu set objek fisik atau abstrak ke dalam kelas-kelas yang serupa. Cluster adalah kumpulan objek data yang memiliki kemiripan satu sama lain dalam cluster yang sama dan berbeda dengan objek di cluster lain [4]

K-means merupakan algoritma yang umum digunakan dalam proses clustering. Dalam algoritma ini mencari sejumlah cluster yang ditentukan dalam hal kedekatan titik data satu sama lain [5].

Hasil penelitian ini diharapkan menjadi referensi bagi Restu Baru Mart untuk meningkatkannya metode dan sistem informasi untuk menentukan pengelompokan barang sehingga dapat digunakan untuk menentukan stok minimum. Selain itu, model pengelompokan produk ini juga diharapkan dapat digunakan sebagai referensi untuk ritel lain pada umumnya, sebagai dukungan untuk berbagai kebutuhan pengambilan keputusan mereka.

METODE

a. Variabel Penelitian

Data yang digunakan pada penelitian ini diperoleh dengan cara mengambil langsung data hasil penjualan di Restu Baru Mart. Variabel penelitian tersebut meliputi nama produk, tanggal transaksi, jumlah persediaan barang dan jumlah penjualan. Periode waktu pengambilan data adalah bulan Juli 2018. Setelah data – data diperoleh, maka data tersebut dimasukkan ke dalam sistem kemudian disimpan sebagai basis pengetahuan sistem.

b. Flowchart Penelitian

Penelitian ini akan melakukan simulasi penentuan cluster barang berdasarkan data penjualan menentukan item mana yang masuk ke cluster laris atau tidak laris. Data akan dikelompokkan ke 3 cluster, cluster nilai centroid tertinggi akan diberi label sebagai grup item yang laris, sedangkan yang terendah nilai centroid akan diberi label sebagai kelompok item yang tidak laris. Proses pengelompokan hanya berlaku untuk item data yang memiliki jumlah transaksi > 0.

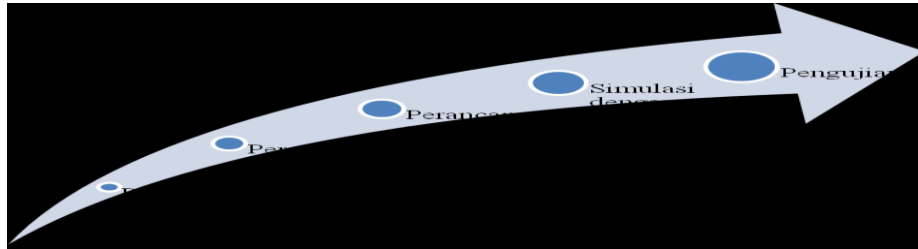
Dalam penelitian ini menggunakan metode K-Mens Clusterring untuk mengelompokkan hasil penjualan. Berikut ini langkah – langkah melakukan perhitungan dengan algoritma *k-means* [6]:

1. Menentukan data yang akan di cluster
2. Menentukan pusat cluster data centroid secara random
3. Menghitung jarak dengan pusat cluster dengan rumus *eucliden* =

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \dots (1)$$
4. Update jarak dimulai dari iterasi awal

5. Menentukan cluster dari iterasi awal hingga iterasi terakhir
6. Lakukan cluster seraca terus menerus hingga data terakhir dengan data sebelumnya bernilai sama

Flowchart dalam penelitian ini dapat dilihat pada gambar 1.

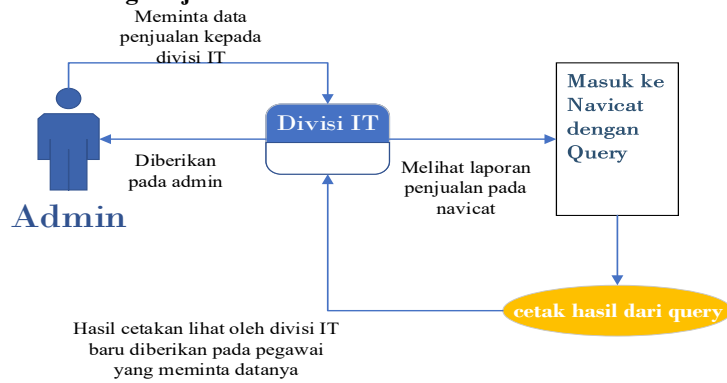


Gambar 1 Tahap Penelitian

Dalam penelitian ini dilakukan beberapa tahap yang meliputi pengumpulan dan analisis data, perancangan algoritma K-Means Clustering, perancangan proses, simulasi dengan Weka dan pengujian.

HASIL

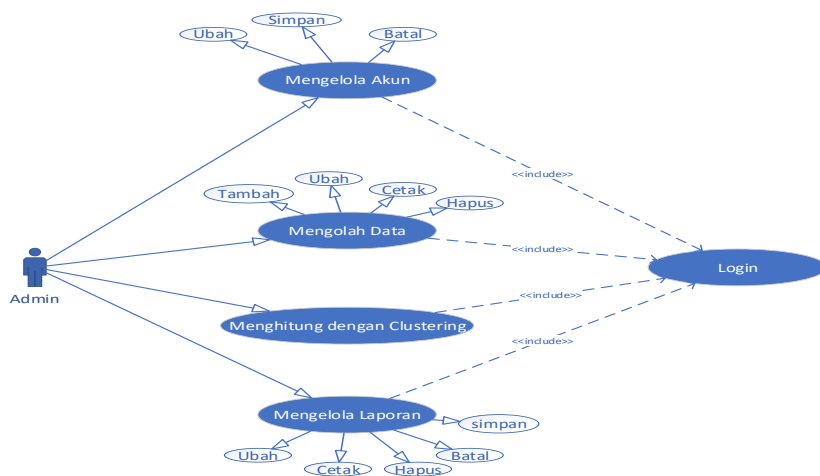
a. Analisa Sistem Yang Berjalan



Gambar 2 Sistem yang Berjalan

b. Use Case Diagram

Digunakan untuk menggambarkan proses apa saja yang dilakukan dari sudut pandang pengguna atau *admin*. Dibawah ini adalah *use case* diagram aplikasi menentukan barang laris, kurang laris dan tidak laris.



Gambar 3 Desain Use Case Diagram

c. Perancangan Algoritma Clustering

Table 1 Data Penjualan

Tgl	Kategori	Nama barang	Persediaan	jml	C
03/Jul/18	Nursing &Feeding	ULTRA CHO KCL KTK 125ML	150	1	
03/Jul/18	Nursing &Feeding	POCARI SWEAT 500ML	150	1	
03/Jul/18	Underwear	CD ANAK CEW XXL	5200	3	
03/Jul/18	Pakaian	LGG LIST MICKEY SMILE AS B	5150	2	C1
03/Jul/18	Underwear	CD CE PRINT 9-10	4100	3	
03/Jul/18	Underwear	CD CE PRINT 9-10	4100	3	
03/Jul/18	Nursing &Feeding	BENDERA BTL STRAW 180ML	25	1	
03/Jul/18	Nursing &Feeding	MIZONE LYCHEE LEMON 500ML	325	1	
03/Jul/18	Underwear	BH SPORT F 003 OSSY,SKIN	2100	2	
03/Jul/18	Pakaian	POPOK KAOS PE	3225	12	
03/Jul/18	Pakaian	SPOROT BJ PISCCO WRN 16	2160	3	C2
03/Jul/18	Bedding	SELIMUT NOVIA 288	5200	6	
03/Jul/18	Aksesories	KERTAS KADO CUSSENS SOAP 75G	180	1	
03/Jul/18	Skin Care	SOFT&SMOOTH	4300	1	
03/Jul/18	Pakaian	BJ YKS WRN PRINT A 601	3180	3	
03/Jul/18	Pakaian	GRITO TALI SABLON	2270	6	C3
03/Jul/18	Aksesories	KK LIPAT/LURUS ISI 3 MTF CMPR	300	1	
03/Jul/18	Underwear	CD SBL GN XL	5400	3	
03/Jul/18	Bathing	HDK DRAGONFLY 55X100	5000	1	

Tabel 2 Pusat Cluster/ Centroid Awal

C1	5150,2
C2	2160,3
C3	2270,6

Tabel 3 Jarak dengan Pusat Cluster Iterasi Pertama

Data ke -i	C1	C2	C3	Cluster
1	5000.00	2010.00	2120.01	2
2	5000.00	2010.00	2120.01	2
3	50.01	3040.00	2930.00	1
4	0.00	2990.00	2880.00	1
5	1050.00	1940.00	1830.00	1
6	1050.00	1940.00	1830.00	1
7	5125.00	2135.00	2245.01	2
8	4825.00	1835.00	1945.01	2
9	3050.00	60.01	170.05	2
10	1925.03	1065.04	955.02	3

11	2990.00	0.00	110.04	2
12	50.16	3040.00	2930.00	1
13	4970.00	1980.00	2090.01	2
14	850.00	2140.00	2030.01	1
15	1970.00	1020.00	910.00	3
16	2880.00	110.04	0.00	3
17	4850.00	1860.00	1970.01	2
18	250.00	3240.00	3130.00	1
19	150.00	2840.00	2730.00	1

Tabel 4 Hasil Clusterring Iterasi Pertama

data ke- i	cluster1		cluster2		cluster3	
	X	y	X	y	X	Y
1	0	0	150	1	0	0
2	0	0	150	1	0	0
3	5200	3	0	0	0	0
4	5150	2	0	0	0	0
5	4100	3	0	0	0	0
6	4100	3	0	0	0	0
7	0	0	25	1	0	0
8	0	0	325	1	0	0
9	0	0	2100	2	0	0
10	0	0	0	0	3225	12
11	0	0	2160	3	0	0
12	5200	6	0	0	0	0
13	0	0	180	1	0	0
14	4300	1	0	0	0	0
15	0	0	0	0	3180	3
16	0	0	0	0	2270	6
17	0	0	300	1	0	0
18	5400	3	0	0	0	0
19	5000	1	0	0	0	0
Jml	38450	22	5390	11	8675	21
countif	8	8	8	8	3	3
Bagi	4806.25	2.75	673.75	1.375	2891.667	7

Tabel 5 Pusat centroid pada iterasi pertama

C1	4806.25,2.75
C2	673.75,1.375
C3	2891.667,7

Tabel 6 Jarak dengan Pusat Cluster Iterasi Kedua

Data ke -i	C1	C2	C3	Cluster
1	4656.25	523.75	2741.67	2
2	4656.25	523.75	2741.67	2
3	393.75	4526.25	2308.34	1
4	343.75	4476.25	2258.34	1
5	706.25	3426.25	1208.34	1
6	706.25	3426.25	1208.34	1
7	4781.25	648.75	2866.67	2
8	4481.25	348.75	2566.67	2
9	2706.25	1426.25	791.68	3
10	1581.28	2551.27	333.37	3
11	2646.25	1486.25	731.68	3
12	393.76	4526.25	2308.33	1
13	4626.25	493.75	2711.67	2
14	506.25	3626.25	1408.35	1
15	1626.25	2506.25	288.36	3
16	2536.25	1596.26	621.67	3
17	4506.25	373.75	2591.67	2
18	593.75	4726.25	2508.34	1
19	193.76	4326.25	2108.34	1

Tabel 7 Hasil Clusterring Iterasi Kedua

data ke-i	cluster1		cluster2		cluster3	
	x	y	X	y	x	Y
1	0	0	150	1	0	0
2	0	0	150	1	0	0
3	5200	3	0	0	0	0
4	5150	2	0	0	0	0
5	4100	3	0	0	0	0
6	4100	3	0	0	0	0
7	0	0	25	1	0	0
8	0	0	325	1	0	0
9	0	0	0	0	2100	2
10	0	0	0	0	3225	12
11	0	0	0	0	2160	3
12	5200	6	0	0	0	0
13	0	0	180	1	0	0
14	4300	1	0	0	0	0
15	0	0	0	0	3180	3
16	0	0	0	0	2270	6

17	0	0	300	1	0	0
18	5400	3	0	0	0	0
19	5000	1	0	0	0	0
jml	38450	22	1130	6	12935	26
countif	8	8	6	6	5	5
bagi	4806.25	2.75	188.3333	1	2587	5.2

Tabel 8 Pusat centroid pada iterasi kedua

C1	4806.25,2.75
C2	188.3333,1
C3	2587,5.2

Tabel 9 Jarak dengan Pusat Cluster Iterasi Ketiga

Data ke -i	C1	C2	C3	Cluster
1	4656.25	38.33	2437.00	2
2	4656.25	38.33	2437.00	2
3	393.75	5011.67	2613.00	1
4	343.75	4961.67	2563.00	1
5	706.25	3911.67	1513.00	1
6	706.25	3911.67	1513.00	1
7	4781.25	163.33	2562.00	2
8	4481.25	136.67	2262.00	2
9	2706.25	1911.67	487.01	3
10	1581.28	3036.69	638.04	3
11	2646.25	1971.67	427.01	3
12	393.76	5011.67	2613.00	1
13	4626.25	8.33	2407.00	2
14	506.25	4111.67	1713.01	1
15	1626.25	2991.67	593.00	3
16	2536.25	2081.67	317.00	3
17	4506.25	111.67	2287.00	2
18	593.75	5211.67	2813.00	1
19	193.76	4811.67	2413.00	1

Tabel 10 Hasil Clusterring Iterasi Kedua

data ke-i	cluster1		cluster2		cluster3	
	x	y	x	Y	x	Y
1	0	0	150	1	0	0

2	0	0	150	1	0	0
3	5200	3	0	0	0	0
4	5150	2	0	0	0	0
5	4100	3	0	0	0	0
6	4100	3	0	0	0	0
7	0	0	25	1	0	0
8	0	0	325	1	0	0
9	0	0	0	0	2100	2
10	0	0	0	0	3225	12
11	0	0	0	0	2160	3
12	5200	6	0	0	0	0
13	0	0	180	1	0	0
14	4300	1	0	0	0	0
15	0	0	0	0	3180	3
16	0	0	0	0	2270	6
17	0	0	300	1	0	0
18	5400	3	0	0	0	0
19	5000	1	0	0	0	0
Jml	38450	22	1130	6	12935	26
countif	8	8	6	6	5	5
bagi	4806.25	2.75	188.3333	1	2587	5.2

Tabel 11 Pusat centroid pada iterasi ketiga

C1	4806.25,2.75
C2	188.3333,1
C3	2587,5.2

Dari hasil perancangan dengan Algoritma K-Means Clustering maka diperoleh hasil cluster 1 (terlaris) ada 8 data, cluster 2 (kurang laring) ada 6 data dan cluster 3 (tidak laris 5 data)

d. Simulasi K-means Clustering

Dari hasil simulasi diperoleh hasil cluster 0 sebanyak 26%, cluster 1 sebanyak 32%, cluster 2 sebanyak 42%

```

Distortion: 2.765874
BIC-Value : -19.12376

Time taken to build model (full training data) : 0 seconds

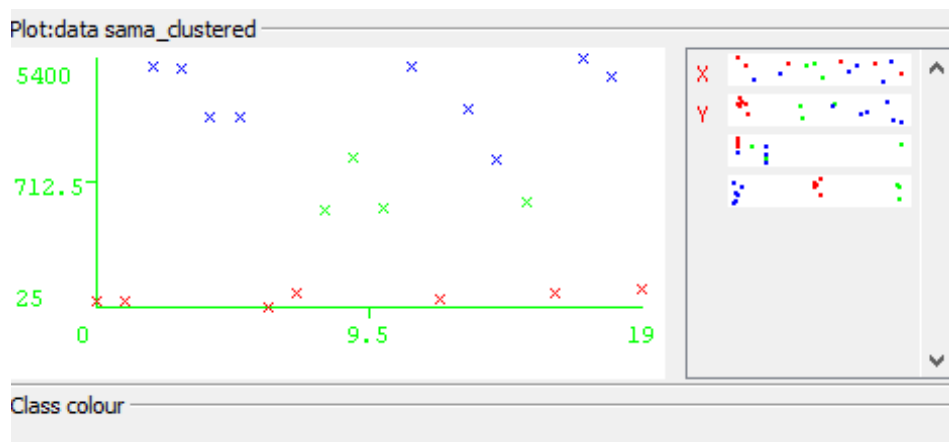
=== Model and evaluation on training set ===

Clustered Instances

0      5 ( 26%)
1      6 ( 32%)
2      8 ( 42%)
    
```

Gambar 4. Hasil 3 cluster dengan data yang sama

Persebaran cluster dapat dilihat pada gambar 5 yang merupakan bentuk visualisasi hasil Clustering.



Gambar 5 Visualisasi cluster

SIMPULAN

Pengelompokan K-Means dapat digunakan dalam proses pengelompokan item ke dalam kategori laris, kurang laris dan tidak laris. Dengan menggunakan data penjualan di Restu Baru Mart pada bulan Juli 2018, ditunjukkan bahwa cluster 1 terdapat 42%, cluster 2 terdapat 32%, cluster 3 terdapat 26%.

DAFTAR PUSTAKA

- [1] Sutrisno, Afriyudi, Widiyanto, "Penerapan data mining pada penjualan menggunakan metode clustering study kasus PT.Indomarco Palembang", Universitas Bina Dharma, Palembang, 2013.
- [2] Giering, M, "Retail sales prediction and item recommendations using customer demographics at store level", SIGKDD Explorations, Volume 10 Issue 2, ACM New York, NY, USA, 10.1145/1540276.1540301. p84-89, 2008.
- [3] Kusriani, "Grouping of Retail Items by Using K-Means Clustering", ISICO, Procedia Computer Science 72, pp.495 – 5027, 2015.
- [4] Phrabu, S., Venatesan, N. "Data Mining And Warehousing, New Age International Publisher", 2007, pp.34-40
- [5] Berry, M.J.A. dan Linoff, G. S., "Data Mining Techniques For Marketing, Sales, and Customer Relationship Management", Second Edition, Wiley Publishing, Inc., Indianapolis, Indiana, 2004,
- [6] Larose, D. T., "Discovering Knowledge in Data: An Introduction to Data Mining". John Wiley and Sons, 2005. pp: 116-126 and 153-158. ISBN: 0471666572. DOI: 10.1002/0471687545