



Computer Based Information System Journal

ISSN (Print): 2337-8794 | E- ISSN : 2621-5292
 web jurnal : <http://ejournal.upbatam.ac.id/index.php/cbis>



ANALISIS *CLUSTERING* MENGGUNAKAN ALGORITMA K-MEANS TERHADAP PENJUALAN PRODUK PADAPT BATAMAS NIAGA JAYA

Mesri Silalahi

Universitas Putera Batam, Indonesia.

INFORMASI ARTIKEL

Diterima Redaksi: 1 September 2018
 Diterbitkan Online: 30 September 2018

KATA KUNCI

Penjualan, Data Mining, K-Means, Rapidminer

KORESPONDENSI

E-mail: mesripb@gmail.com

ABSTRACT

PT Batamas Niaga Jaya was established in 1994 located in Sagulung Mas Indah complex, Sagulung Batu Aji district, Batam. This PT rents out 24 kiosks and 48 counters, in the retail sector, PT has 4 stores, namely BIG supermarket, Department store, Stationary and 7,500 convenience store and 10,000 department stores. 7,500 Convenience Stores and 10,000 Stores PT Batamas Niaga Jaya sells various products including food, beverages, kitchen utensils and equipment, and various daily necessities. Sales transactions for 7,500 convenience stores and 10,000 convenience stores were used as objects for analysis. PT Batamas Niaga Jaya often has problems in product management, the company is unable to classify the best-selling products and is not in demand, so it often lacks selling products and piles of non-salable products. The data used is sales data for the last 3 months (February 2018-April 2018). The data cluster technique uses the K-Means algorithm where the data are grouped into several clusters based on the similarity of the data. The results of the study obtained 2 product groups, namely products that were not in demand with the C0 group and the best-selling product group with the C1 group. The C0 product group that is not in demand consists of 657 products and the C1 group consists of 70 products which are in demand.

I. Latar Belakang

PT Batamas Niaga Jaya berdiri sejak tahun 1994 terletak di komplek Sagulung Mas Indah kecamatan Sagulung Batu Aji Batam. Selain retail, PT Batamas Niaga Jaya juga menyewakan kios counter. PT ini menyewakan 24 kios dan 48 counter, sedangkan di bidang retail PT ini memiliki 4 toko retail yaitu BIG supermarket, Departement store,

Stationary serta Toko Serba 7.500 dan Toko serba 10.000. Khusus di bidang retail Toko Serba 7.500 dan Toko Serba 10.000 PT Batamas Niaga Jaya menjual berbagai macam produk diantaranya produk-produk makanan, minuman, peralatan dan perlengkapan dapur, dan masih banyak lagi produk yang dijual untuk kebutuhan sehari-hari.

Dalam proses penjualan PT Batamas Niaga Jaya sering mengalami

kendala khususnya dalam pengelolaan produk, perusahaan ini kurang dapat mengelompokkan produk yang laris dan yang tidak laris terjual. Sehingga kesulitan yang dialami yaitu seringnya kekurangan stok produk yang laku karena penjualannya tinggi. Dan menumpuknya produk yang tidak laku di gudang karena penjualannya rendah.

Dalam penelitian ini, pemilihan data transaksi penjualan barang Toko Serba 7.500 dan Toko Serba 10.000 PT Batamas Niaga Jaya dijadikan objek untuk dianalisis dikarenakan dalam data transaksi penjualan barang tersimpan potensi – potensi atau fakta yang didokumentasikan untuk digali dan diproses menjadi informasi yang bermanfaat bagi perusahaan agar nantinya dapat mengurangi permasalahan dalam pengelompokan produk sehingga kekurangan dan kelebihan stok dapat dihindari serta dapat membantu pihak manajemen dalam pengambilan keputusan seperti pengadaan program promo untuk produk-produk tertentu. Salah satu teknik *cluster* data adalah dengan menggunakan algoritma K-Means. Dengan menggunakan metode ini, data-data yang telah didapatkan dapat dikelompokkan kedalam beberapa *cluster* berdasarkan kemiripan dari data-data tersebut, sehingga data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* dan yang memiliki karakteristik yang berbeda dikelompokkan dalam *cluster* yang lain yang memiliki karakteristik yang sama. Tujuan dari penelitian ini yaitu: Untuk menganalisa data penjualan dengan metode algoritma K-Means yang dapat membantu mengelompokkan berbagai produk pada PT Batamas Niaga Jaya.

II. Kajian Literatur Produk

Menurut Kotler [1], produk adalah apa saja yang dapat ditawarkan ke pasar untuk diperhatikan, diperoleh, digunakan atau dikonsumsi yang dapat memenuhi kebutuhan atau keinginan konsumen. Menurut Fandy Tjiptono Produk potensial, yaitu segala macam tambahan dan perubahan yang mungkin dikembangkan untuk suatu produk dimasa datang. Sementara itu menurut Levitt dalam buku “*The Marketing Imagination The potensial product*” Produk potensial merupakan tampilan dan manfaat tambahan yang berguna bagi pelanggan atau mungkin menambah kepuasan pelanggan.

Pendapat [2] menjelaskan, produk merupakan pemuas kebutuhan yang ditawarkan oleh perusahaan dan mencakup produk fisik dan jasa atau dapat juga merupakan gabungan keduanya (Perreault dan McCarthy, 2009: 31). Produk mencakup segala sesuatu yang dapat ditawarkan ke pasar untuk diperhatikan, dimiliki, digunakan atau dikonsumsi sehingga dapat memuaskan kebutuhan atau keinginan. Produk mencakup barang, jasa, pengalaman, acara, orang, tempat, organisasi, informasi dan gagasan (Kotler 2008: 4). Uraian di atas menyimpulkan bahwa pelanggan juga membeli kepuasan, bukan produk fisik atau jasa saja. Produk harus dapat dilihat dari sudut pandang pelanggan dan bagaimana pelanggan berpikir bahwa produk dapat sesuai dengan tujuan pembeliannya.

Menurut Tjiptono (2008:95) dalam [3], Produk merupakan segala sesuatu yang dapat ditawarkan produsen untuk diperhatikan, dicari, dibeli, digunakan atau dikonsumsi pasar sebagai

memenuhi kebutuhan atau keinginan pasar yang bersangkutan. Produk yang ditawarkan meliputi barang fisik seperti sepeda motor, computer, televisi sedangkan produk jasa meliputi restoran, penginapan dan transportasi. Produk bisa berupa manfaat tangible maupun intangible yang dapat memuaskan pelanggan. Produk secara konseptual, yaitu pemahaman subyektif dari produsen atas sesuatu yang bisa ditawarkan sebagai usaha untuk mencapai tujuan organisasi melalui pemenuhan kebutuhan dan keinginan konsumen sesuai dengan kompetensi dan kapasitas organisasi serta daya beli pasar.

Penjualan

Menurut Pride dan Ferrell (McLeod:2001:449) penjualan adalah kegiatan perorangan dan organisasi yang memudahkan dan mempercepat hubungan pertukaran yang memuaskan dalam lingkungan yang dinamis melalui penciptaan, pendistribusian, promosi dan penentuan harga barang, jasa dan gagasan. Sementara Fuad (2005:119) menyatakan bahwa penjualan adalah proses perpindahan barang dan atau jasa dari produsen ke konsumen atau semua kegiatan yang berhubungan dengan arus barang dan atau jasa dari produsen ke konsumen. Menurut Sidharta (1995:46), Penjualan adalah struktur interaksi antara manusia, peralatan, metode-metode dan kontrol-kontrol yang disusun untuk mencapai tujuan-tujuan tertentu. Prosedur penjualan melibatkan beberapa bagian dalam perusahaan dengan maksud agar penjualan yang terjadi dapat diawasi dengan baik. Bagian-bagian yang terkait dengan prosedur penjualan adalah bagian pesanan penjualan, bagian kredit, bagian gudang, bagian pengiriman dan bagian

penagihan. Masing-masing bagian memiliki peranan dalam menjalankan prosedur penjualan [4].

Menurut Assauri (2007:121-125), kegiatan penjualan merupakan kegiatan pelengkap atau suplemen dari pembelian untuk memungkinkan terjadinya transaksi. Jadi kegiatan pembelian dan penjualan merupakan satu kesatuan untuk dapat terlaksananya transfer hak atau transaksi. Oleh karena itu, kegiatan penjualan terdiri dari serangkaian kegiatan yang meliputi penciptaan permintaan (*demand*), menemukan pembeli, negosiasi harga dan syarat-syarat pembayaran. Dalam hal penjualan ini, penjual harus menentukan kebijaksanaan dan prosedur yang akan diikuti untuk memungkinkan dilaksanakannya rencana penjualan yang telah ditetapkan. Walaupun sering terlihat bahwa usaha atau inisiatif mencari produk datangnya dari pembeli, tetapi penjual haruslah berusaha untuk dapat menemukan pembeli agar transaksi penjualan dapat terlaksana [5].

Data Mining

Menurut Widodo (2013:1) [6] Data mining adalah analisa terhadap data untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut. Secara garis besar, data mining dapat dikelompokkan menjadi 2 kategori utama, yaitu:

1. *Descriptive mining*, yaitu proses untuk menemukan karakteristik penting dari data dalam satu basis data. Teknik data mining yang termasuk *descriptive mining* adalah *cluster ing*, *asosiasi*, dan *sequential mining*.

2. *Predictive*, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variable lain di masa depan. Salah satu teknik yang terdapat dalam predictive mining adalah klasifikasi.

Secara sederhana data mining biasa dikatakan sebagai proses penyaring atau “menambang” pengetahuan dari sejumlah data yang besar. Istilah lain untuk data mining adalah *Knowledge Discovery in Database* (KDD). Walaupun data mining sendiri adalah bagian dari tahapan proses KDD seperti yang terlihat pada Gambar 1.

1. *Data Selection*

Menciptakan himpunan data target, pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (discovery) akan dilakukan. Hasil seleksi disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. *Pre-processing / Cleaning*

Pre-processing dan cleaning data merupakan operasi dasar yang dilakukan seperti penghapusan noise. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Data bisa diperkaya dengan data atau informasi eksternal yang relevan.

3. *Transformation*

Merupakan proses integrasi pada data yang telah dipilih, sehingga data sesuai untuk proses data mining. Merupakan proses yang sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data mining*

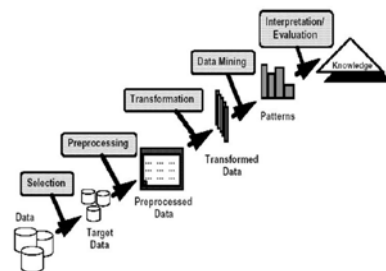
Pemilihan tugas data mining merupakan pemilihan goal dari proses KDD misalnya karakterisasi, klasifikasi, regresi, *clustering*, asosiasi, dan lain-lain. Pemilihan tugas data mining merupakan pemilihan goal dari proses KDD misalnya karakterisasi, klasifikasi, regresi, *clustering*, asosiasi, dan lain-lain. Pemilihan teknik, metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation/ Evaluation*

Yaitu penerjemahan pola-pola yang dihasilkan dari data mining. Pola informasi yang dihasilkan perlu ditampilkan dalam bentuk yang mudah dimengerti. Tahap ini melakukan pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

Tujuan dari data mining yaitu:

- Explonatory, yaitu untuk menjelaskan beberapa kegiatan opservasi atau kondisi.
- Confirmatory, yaitu untuk mengkonfirmasi suatu hipotesis yang telah ada.
- Exploratory, yaitu untuk menganalisis data baru suatu relasi yang janggal.



Gambar 1. KDD

Data Mining sangat perlu dilakukan terutama dalam mengelola data yang sangat besar untuk memudahkan aktivitas *recording* atau perekaman suatu transaksi dan untuk proses Data Warehousing agar dapat memberikan informasi yang akurat bagi pengguna Data Mining [7].

Dalam jurnal [8], Tugas data mining adalah Deskriptif (menemukan pola data) dan prediktif (meramalkan perilaku data yang tersedia berdasarkan model). Data mining melakukan eksekusi berdasarkan teknik komputasi dan metode dari pembelajaran mesin statistik dan pengenalan pola. Teknik data mining menavigasi sejumlah besar database dan mengekstrak sejumlah data yang akan diterjemahkan dan ditulis menjadi informasi yang berguna sesuai kebutuhan. Proses data mining dapat dicirikan sebagai proses berulang multi-tahap yang melibatkan seleksi data, pembersihan data, penerapan algoritma data mining dan evaluasi.

Fungsi data mining meliputi klasifikasi, *clustering*, analisis asosiasi, analisis time series, dan analisis outlier. (i) Klasifikasi adalah proses menemukan satu set model atau fungsi yang menjelaskan dan membedakan kelas data atau konsep, untuk tujuan memprediksi kelas dari objek yang tidak diketahui. (ii) *Clustering*, analisis objek data tanpa model kelas diketahui. (iii) analisis Asosiasi adalah penemuan aturan asosiasi menampilkan kondisi atribut-nilai yang frequently terjadi bersama-sama dalam suatu himpunan data. (iv) analisis time series terdiri dari metode dan memakai teknik untuk menganalisis data time series untuk mengekstrak statistik bermakna dan karakteristik lain dari data. (v) analisis outlier menjelaskan dan

model keteraturan atau tren untuk objek yang perilakunya berubah dari waktu ke waktu [9].

Clustering

Data mentah yang sangat besar sulit untuk dianalisis dan dipahami, oleh karenanya perlu adanya pengelompokan/ *clustering*. Dalam hal ini pengelompokan bertujuan untuk meningkatkan pemahaman terhadap data, untuk menilai kualitas dari data tersebut. K-means algoritma *clustering* merupakan adalah alat pengelompokan yang paling populer digunakan dalam aplikasi ilmiah dan industri. K mewakili masing-masing *cluster* k oleh mean (atau rata-rata tertimbang) dari poin-nya, yang disebut massa. Pusat massa dari *cluster* adalah titik koordinat rata-rata dari koordinat semua titik dalam *cluster* [10].

Clustering merupakan salah satu teknik dari salah satu fungsionalitas data mining, algoritma *cluster ing* merupakan algoritma pengelompokan sejumlah data menjadi kelompok-kelompok data tertentu (*cluster*). Objek data yang terletak di dalam satu *cluster* harus mempunyai kemiripan. Sedangkan yang tidak berada didalam satu *cluster* tidak mempunyai kemiripan. Setiap *cluster* memiliki *centroid* yang merupakan suatu besaran yang dihitung dari rata-rata nilai tiap *items* dari suatu *cluster* dan juga memiliki *medoid* yang merupakan item yang letaknya paling tengah. Sementara jarak antar *cluster* didefinisikan dengan menggunakan beberapa metode-metode untuk menentukan *cluster* mana yang berdekatan.

1. *Single link*: jarak terkecil antara satu elemen dalam suatu *cluster* dengan elemen dalam suatu *cluster* dengan

- elemen lain di *cluster* yang berbeda.
2. *Complete link*: jarak terbesar antara satu elemen dalam suatu *cluster* dengan elemen lain di *cluster* yang berbeda.
 3. *Average*: jarak rata-rata antara satu elemen dalam suatu *cluster* dengan elemen lain di *cluster* yang berbeda.
 4. *Centroid*: jarak antara *centroid* dari tiap *cluster* dengan *centroidcluster* lainnya
 5. *Medoid*: jarak antara *medoid* dari tiap *cluster* dengan *medoidcluster* lainnya.

Clustering memiliki banyak algoritma dengan beberapa kategori yaitu *hierarchical* yang menentukan sendiri jumlah *cluster* yang dihasilkan misalnya algoritma SNN, COBWEB, Chandeon dan Rock. *Partitional* yang mengelompokkan data kedalam *k cluster* dimana *k* adalah banyak *cluster* dari input user misalnya algoritma CLARA, K-Means, EM dan Bond Energy. Serta *cluster ing large data* yang memiliki volume data paling besar misalnya algoritma BIRCH, DBScan dan Cactus [11].

Algoritma K-means

K-means [12] merupakan salah satu metode data *cluster ing* non-hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster i* kelompok. Metode ini mempartisi ke dalam *cluster i* kelompok sehingga data yang memiliki karakteristik yang sama (*high intra-class similarity*) dikelompokkan ke dalam satu *cluster* yang sama dan yang memiliki karakteristik yang berbeda (*low Inter-class similarity*) dikelompokkan pada

kelompok yang lain. Proses *cluster ing* dimulai dengan mengidentifikasi data yang akan di *cluster* , $x_{ij}(i=1,\dots,n;j=1,\dots,m)$ dengan *n* adalah jumlah data yang akan di *cluster* dan *m* adalah jumlah variabel. Pada awal iterasi, pusat setiap *cluster* ditetapkan secara bebas (sembarang), $(k=1,\dots,=1,\dots,m)$. Kemudian dihitung jarak antara setiap data dengan setiap pusat *cluster* .

Algoritma dasar dalam *k-means* adalah:

1. Tentukan jumlah *cluster* (*k*), tetapkan pusat *cluster* sembarang.
2. Hitung jarak setiap data ke pusat *cluster* menggunakan persamaan:

$$d_{ik} = \sqrt{\sum_{j=1}^m \{x_{ij} - c_{kj}\}^2}$$

3. Kelompokkan data ke dalam *cluster* yang dengan jarak yang paling pendek menggunakan persamaan

$$\min \sum_{k=1}^k d_{ik}$$

4. Hitung pusat *cluster* yang baru menggunakan persamaan

$$c_{kj} = \frac{1}{p} \sum_{i=1}^p x_{ij}$$

Langkah 2 sampai dengan 4 akan diulangi sehingga tidak ada lagi data yang berpindah ke *cluster* yang lain. Prinsip utama pengelompokkan dengan *Kmeans* adalah meminimumkan nilai jarak elemen-elemen dengan titik pusat (*centroid*) dalam *cluster* [13].

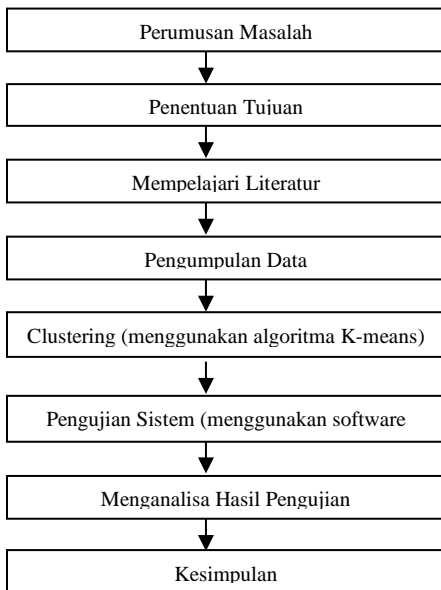
Rapidminer

RapidMiner merupakan software/perangkat lunak untuk pengolahan data, dengan menggunakan prinsip dan algoritma data mining, Rapidminer mengekstrak pola-pola dari data set yang besar dengan

mengkombinasikan metode statistika, kecerdasan buatan dan database.

Rapidminer memudahkan penggunaanya dalam melakukan perhitungan data yang sangat banyak dengan menggunakan operator-operator. Operator ini berfungsi untuk memodifikasi data. Data dihubungkan dengan node-node pada operator kemudian pengguna hanya tinggal menghubungkannya ke node hasil untuk melihat hasilnya. Hasil yang diperlihatkan RapidMiner pun dapat ditampilkan secara visual dengan grafik menjadikan RapidMiner adalah salah satu software pilihan untuk melakukan ekstraksi data dengan metode-metode data mining [14].

Kerangka Pemikiran

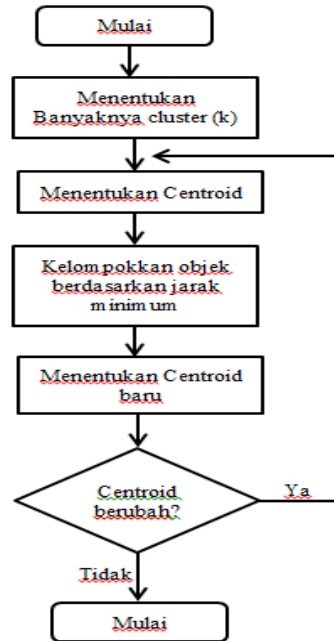


Gambar 2. Kerangka Pemikiran

III. Metodologi Desain Penelitian

Untuk mengelompokkan produk PT Batamas Niaga Jaya yang laris dan

tidak laris dengan algoritma k-means maka secara umum alur yang digunakan seperti gambar di bawah ini:



Gambar 3. Desain Penelitian

Tahapan-tahapan pada desain penelitian dapat dijelaskan seperti berikut ini :

1. Menentukan banyaknya *cluster* (k)
Untuk menganalisis data penjualan Toko Serba 7.500 dan Toko Serba 10.000 PT Batamas Niaga Jaya dengan algoritma k-means dimulai dengan penentuan jumlah *cluster* (k). Penentuan jumlah *cluster* dilakukan dengan pemilihan secara acak.
2. Menentukan centroid
Setelah jumlah *cluster* ditentukan selanjutnya dalam tahapan ini ditetapkan nilai k secara random. Nilai k ditahap ini sekaligus dijadikan menjadi pusat dari *cluster* awal atau centroid atau disebut mean/means.

3. Menentukan objek berdasarkan jarak minimum
Setelah nilai centroid awal ditentukan maka dihitung jarak setiap data yang ada terhadap terhadap masing-masing centroid menggunakan rumus Euclidian seperti yang dijelaskan pada bab 2. Setelah jarak setiap data terhadap centroid diperoleh maka selanjutnya data dikelompokkan berdasarkan jarak minimum kemudian klasifikasikan data berdasarkan kedekatannya dengan centroid.
4. Menentukan centroid baru
Apabila nilai centroid masih berubah maka tahapan diulang dengan menentukan centroid baru sampai centroid tidak berubah atau stabil.

Lokasi Penelitian

PT Batamas Niaga Jaya. Perusahaan ini telah berdiri sejak tahun 1994 terletak di komplek Sagulung Mas Indah kecamatan Sagulung Batu Aji Batam. Selain retail, PT Batamas Niaga Jaya juga menyewakan kios counter. PT ini menyewakan 24 kios dan 48 counter, sedangkan di bidang retail PT ini memiliki 4 toko retail yaitu BIG supermarket, Departement store, Stationary serta Toko Serba 7.500 dan Toko serba 10.000. Khusus di bidang retail, PT Batamas Niaga Jaya menjual berbagai macam produk diantaranya produk-produk makanan, minuman, peralatan dan perlengkapan dapur, dan masih banyak lagi produk yang dijual untuk kebutuhan sehari-hari.

Analisis Data

Data yang digunakan untuk penelitian ini berupa data set penjualan

usaha retail Toko Serba 7.500 dan Toko Serba 10.000 mulai dari bulan Februari 2018-April 2018 yang dikelompokkan menjadi 75 kategori seperti tabel 1 di bawah.

Tabel 1. Data Set Penjualan Barang Berdasarkan Kategori Produk

No	Kategori	Jumlah Item (Jenis)	Jumlah Penjualan (Unit)		
			Februari	Maret	April
1	Baskom	5	119	63	68
2	Bola	6	0	0	73
3	Baterai	6	82	48	0
4	Botol	4	108	64	31
5	Box	2	18	8	2
6	Cermin	1	106	9	0
7	Clean Duster	8	69	71	16
8	Corong	5	10	10	11
9	Cutter	107	57	77	392
10	Cleaner Wc	3	21	14	15
11	Food	1	35	32	0
12	Gantungan	9	79	47	71
13	Garpu	4	12	10	25
14	Gayung	5	76	76	57
15	Gelas	36	356	154	183
16	Handuk	1	0	0	6
17	Hanger	7	137	98	131
18	Hook	9	57	55	57
19	Kain	5	167	71	122
20	Kemoceng	1	29	12	21
21	Keranjang	26	362	181	59
22	Kran	3	27	11	5
23	Lap	2	4	2	9
24	Lem	5	0	0	38
25	Mainan	57	2,271	1,322	1,291
26	Mangkok	14	142	54	61
27	Nampan	8	289	69	92
28	Obeng	4	14	16	22
29	Parutan	3	37	22	23
30	Payung	1	9	8	12

Analisis Data Model

Sesuai dengan dataset yang sudah dikelompokkan berdasarkan

kategori seperti pada tabel 1 dimana masing-masing kategori terdiri dari beberapa item yaitu sebanyak 727 item. Data set penjualan ini memiliki beberapa atribut yaitu no, nama barang, kode, barcode, qty jual, nominal, modal dan profit. Namun dalam penelitian ini tidak semua atribut digunakan, dalam penelitian ini atribut data penjualan yang dianalisis hanya qty jual per bulan dari setiap item saja (data yang ditampilkan hanya sebagai contoh untuk data lengkap akan dilampirkan).

Tabel 2. Analisis Data Model

No	Nama Barang	Jumlah Penjualan		
		Februari	Maret	April
1	0029 alas panas motif	0	0	4
2	3273 alas panas motif	0	0	1
3	393#15*15 alas panas persegi	0	0	1
4	539 alas panas circle	0	0	1
5	alas kaki	0	0	30
6	2's qiangshun perekat tikus	0	0	0
7	items fresh non barcode	1	0	2
8	02 shunfeng baskom	58	31	29
9	057 centa t.sendok	7	5	9
10	6884-baskom	36	16	17
11	802 ke wei baskom	7	1	2
12	8032 baskom	11	7	11
13	3s eveready s91 bp20 sil 3	3	1	0
14	3s eveready s92 bp20 sil 4	1	0	0
15	4's eveready 1012 bp20 aaa mrh 1	16	9	0
16	4's eveready 1015 bp20 aa mrh kcl 2	30	11	0
17	energizer e91 aa bp2+1 max 1	16	20	0
18	energizer e92 aaa bp2+1max 2	16	7	0
19	1199 bola sport	0	0	6
20	7788 bola smiley	0	0	2
21	l688-12 bola kecil	0	0	2

22	lf189 bola karet	0	0	52
23	v010 bola emo kuning besar	0	0	4
24	v014 bola kartun	0	0	7
25	059 tmpm minum	50	32	10
26	1189 easy press juice	35	15	2
27	603-botol semprot	2	1	5
28	a-6/ec-1386 btl kcp	21	16	14
29	8's rzk mancis	15	7	0
30	tempat sabun tempel	3	1	2

IV. Pembahasan

Implementasi Algoritma K-Means

a. Penetapan jumlah cluster (k)

Penetapan jumlah cluster (K) dalam penelitian ini dilakukan dengan dipilih secara random sebanyak 2 cluster .

Tabel 3. Titik Awal Tiap Cluster

Titik Pusat Awal	Februari (Fe)	Maret (Ma)	April (AP)
Cluster 0	154	8	40
Cluster 1	133	56	84

b. Menentukan centroid dan Jarak Terpendek

Jarak data data penjualan ke-1 ke pusat cluster :

$$C(0) = \sqrt{(0-154)^2 + (0-8)^2 + (4-40)^2}$$

$$= \sqrt{23.716 + 64 + 1.296}$$

$$= \sqrt{25.076}$$

$$= 158,3540337$$

$$C(1) = \sqrt{(0-133)^2 + (0-56)^2 + (4-84)^2}$$

$$= \sqrt{17.689 + 3.136 + 6.400}$$

$$= \sqrt{27.225}$$

$$= 165$$

Dari hasil perhitungan di atas diperoleh bahwa jarak data penjualan ke-1 dengan pusat cluster pertama (C0) adalah 158,3540337 serta dengan pusat cluster kedua (C1) adalah 165. Oleh karena itu jarak yang paling dekat (jarak terpendek) dengan pusat cluster adalah 158,3540337.

Jarak data data penjualan ke-2 ke pusat

cluster :

$$\begin{aligned} C(0) &= \sqrt{(0-154)^2+(0-8)^2+(1-40)^2} \\ &= \sqrt{23.716 +64+1.521} \\ &= \sqrt{25.301} \\ &= 159,0628806 \end{aligned}$$

$$\begin{aligned} C(1) &= \sqrt{(0-133)^2+(0-56)^2+(1-84)^2} \\ &= \sqrt{17.689 +3.136+6.889} \\ &= \sqrt{27.714} \\ &= 166,4752233 \end{aligned}$$

Dari hasil perhitungan di atas diperoleh bahwa jarak data penjualan ke-2 dengan pusat *cluster* pertama (C0) adalah 159,0628806 serta dengan pusat *cluster* kedua (C1) adalah 166,4752233. Oleh karena itu jarak yang paling dekat (jarak terpendek) dengan pusat *cluster* adalah 159,0628806.

Jarak data data penjualan ke-3 ke pusat *cluster* :

$$\begin{aligned} C(0) &= \sqrt{(0-154)^2+(0-8)^2+(1-40)^2} \\ &= \sqrt{23.716 +64+1.521} \\ &= \sqrt{25.301} \\ &= 159,0628806 \end{aligned}$$

$$\begin{aligned} C(1) &= \sqrt{(0-133)^2+(0-56)^2+(1-84)^2} \\ &= \sqrt{17.689 +3.136+6.889} \\ &= \sqrt{27.714} \\ &= 166,4752233 \end{aligned}$$

Dari hasil perhitungan di atas diperoleh bahwa jarak data penjualan ke-3 dengan pusat *cluster* pertama (C0) adalah 159,0628806 serta dengan pusat *cluster* kedua (C1) adalah 166,4752233. Oleh karena itu jarak yang paling dekat (jarak terpendek) dengan pusat *cluster* adalah 159,0628806.

Jarak data data penjualan ke-4 ke pusat *cluster* :

$$\begin{aligned} C(0) &= \sqrt{(0-154)^2+(0-8)^2+(1-40)^2} \\ &= \sqrt{23.716 +64+1.521} \\ &= \sqrt{25.301} \\ &= 159,0628806 \end{aligned}$$

$$\begin{aligned} C(1) &= \sqrt{(0-133)^2+(0-56)^2+(1-84)^2} \\ &= \sqrt{17.689 +3.136+6.889} \\ &= \sqrt{27.714} \end{aligned}$$

$$= 166,4752233$$

Dari hasil perhitungan di atas diperoleh bahwa jarak data penjualan ke-4 dengan pusat *cluster* pertama (C0) adalah 159,0628806 serta dengan pusat *cluster* kedua (C1) adalah 166,4752233. Oleh karena itu jarak yang paling dekat (jarak terpendek) dengan pusat *cluster* adalah 159,0628806.

Jarak data data penjualan ke-5 ke pusat *cluster* :

$$\begin{aligned} C(0) &= \sqrt{(0-154)^2+(0-8)^2+(30-40)^2} \\ &= \sqrt{23.716 +64+100} \\ &= \sqrt{23.880} \\ &= 154,5315501 \end{aligned}$$

$$\begin{aligned} C(1) &= \sqrt{(0-133)^2+(0-56)^2+(30-84)^2} \\ &= \sqrt{17.689 +3.136+2.916} \\ &= \sqrt{23.741} \\ &= 154,0811474 \end{aligned}$$

Dari hasil perhitungan di atas diperoleh bahwa jarak data penjualan ke-5 dengan pusat *cluster* pertama (C0) adalah 154,5315501 serta dengan pusat *cluster* kedua (C1) adalah 154,0811474. Oleh karena itu jarak yang paling dekat (jarak terpendek) dengan pusat *cluster* adalah 154,0811474.

Demikian juga dengan data penjualan yang lainnya dilakukan perhitungan dengan cara yang sama. Dalam penelitian ini untuk perhitungannya menggunakan Microsoft excel. Hasil perhitungan dan penentuan jarak terpendek dapat dilihat pada tabel di bawah.

Tabel 4. Perhitungan Iterasi ke-1

No	nama barang	Fe	Ma	A p	C0	C1	Jarak terpendek
1	0029 alas panas motif	0	0	4	158.35 403	165	158.35 403
2	3273 alas panas motif	0	0	1	159.06 288	166.47 5223	159.06 288
3	393#15*1 5 alas panas persegi	0	0	1	159.06 288	166.47 5223	159.06 288
4	539 alas panas circle	0	0	1	159.06 288	166.47 5223	159.06 288
5	alas kaki	0	0	3 0	154.53 155	154.08 1147	154.08 115
6	2's qiangshun perekat tikus	0	0	0	159.31 102	166.97 6046	159.31 102
7	items fresh non barcode	1	0	2	157.85 12	165.17 8691	157.85 12
8	02 shunfeng baskom	58	31	2 9	99.327 74	96.306 8014	96.306 801
9	057 centa t.sendok	7	5	9	150.26 31	155.24 8188	150.26 31
10	6884-baskom	36	16	1 7	120.48 651	124.49 0964	120.48 651
11	802 ke wei baskom	7	1	2	151.99 342	160.07 8106	151.99 342
12	8032 baskom	11	7	1 1	145.91 436	150.37 952	145.91 436
13	3s eveready s91 bp20 sil 3	3	1	0	156.36 496	164.25 8942	156.36 496
14	3s eveready s92 bp20 sil 4	1	0	0	158.34 456	166.18 0625	158.34 456
15	4's eveready 1012 bp20 aaa mrh 1	16	9	0	143.68 368	151.50 5775	143.68 368
16	4's eveready 1015 bp20 aa mrh kcl 2	30	11	0	130.32 651	140.32 106	130.32 651
17	energizer e91 aa bp2+1 max 1	16	20	0	144.18 044	148.46 2116	144.18 044
18	energizer e92 aaa bp2+1max 2	16	7	0	143.68 368	152.13 8095	143.68 368
19	1199 bola sport	0	0	6	157.91 137	164.03 9629	157.91 137
20	7788 bola smiley	0	0	2	158.82 065	165.97 8914	158.82 065

c. Pengelompokan data berdasarkan *cluster*

Setelah semua data dihitung dan jarak terpendek ditentukan maka selanjutnya yaitu mengelompokkan data ke dalam

cluster masing-masing dimana jarak terpendek diberikan angka 1 selain itu 0. Pengelompokan data untuk perhitungan data di atas seperti tabel 5 di bawah.

Tabel 5. Kelompok Data ke – 1

NO	C0	C1
1	0	1
2	0	1
3	0	1
4	0	1
5	1	0
6	0	1
7	0	1
8	1	0
9	0	1
10	0	1
11	0	1
12	0	1
13	0	1
14	0	1
15	0	1
16	0	1
17	0	1
18	0	1
19	0	1
20	0	1

d. Penentuan Pusat *Cluster* Baru

Setelah diperoleh anggota dari tiap-tiap *cluster* seperti pada tabel 4.3, selanjutnya yaitu menghitung pusat *cluster* baru dengan membagi total jumlah penjualan pada *cluster* yang sama dengan total *cluster*. Pada pengelompokan data 1 total C0 = 674, C1=53, total penjualan untuk *cluster* C0 february =6.360, C0 Maret = 2.803, C0 April = 3.365 sedangkan untuk C1 Februari = 1.490, C1 Maret = 1.915, C1 April = 1.281. Pusat *cluster* baru akan

dihitung sesuai dengan rumus yang sudah dijelaskan pada bab 2 dan yang dihitung sebagai berikut:

$$C0(fe) = 6.360/674 = 9.43620178$$

$$C0(Ma) = 2.803 / 674 = 4.158753709$$

$$C0(AP) = 3.365 / 674 = 4.992581602$$

$$C1(Feb) = 1.490 / 53 = 28.11320755$$

$$C1(Ma) = 1.915 / 53 = 36.13207547$$

$$C1(AP) = 1.281 / 53 = 24.16981132$$

Dari perhitungan di atas diperoleh pusat *cluster* baru seperti pada tabel 6.

Tabel 6. Titik Pusat *Cluster* Baru

Titik Pusat Awal	Februari	Maret	April
Cluster 0	9.43620178	4.158753709	4.992581602
Cluster 1	28.11320755	36.13207547	24.16981132

Dikarenakan pusat *cluster* yang baru tidak sama dengan pusat *cluster* awal maka proses akan diulang kembali mulai dari menentukan centroid dengan proses sama seperti sebelumnya.

Tabel 7. Perhitungan dan pengelompokan data Iterasi ke-2

No	Nama Barang	Fe	Ma	A P	C1	C2	arak erpend k
1	0029 alas panas motif	0	0	4	10.35965	50.026999	0.3596
2	3273 alas panas motif	0	0	1	11.057931	51.310033	1.0579
3	393#15*15 alas panas persegi	0	0	1	11.057931	51.310033	1.0579
4	539 alas panas circle	0	0	1	11.057931	51.310033	1.0579
5	alas kaki	0	0	30	27.050104	46.150519	7.0501
6	2s qiangshun perekat tikus	0	0	0	11.457007	51.769287	1.4570
7	items fresh non barcode	1	0	2	9.8701711	50.320507	8.7017
8	02 shunfeng baskom	58	31	29	60.458673	30.706502	0.7065
9	057 centa t.sendok	7	5	9	4.7646801	40.559793	7.6468
10	6884-baskom	36	16	17	31.46472	22.779556	2.7795
11	802 ke wei baskom	7	1	2	4.9868175	46.599687	9.8681
12	8032 baskom	11	7	11	6.8269482	36.262703	8.2694
13	3s eveready s91 bp20 sil 3	3	1	0	8.7366063	49.488541	7.3660
14	3s eveready s92 bp20 sil 4	1	0	0	10.648502	51.233121	0.6485
15	4s eveready 1012 bp20 aaa mfh 1	16	9	0	9.5627916	38.302207	5.6279
16	4s eveready 1015 bp20 aa mfh kcl 2	30	11	0	22.239567	34.91935	2.2395
17	energizer e91 aa bp2+1 max	16	20	0	17.859295	31.482589	7.8592

	1						
18	energizer e92 aaa bp2+1max 2	16	7	0	8.7224996	39.744023	7.2249
19	1199 bola sport	0	0	6	10.361082	49.254658	0.3610
20	7788 bola smilev	0	0	2	10.737443	50.866294	0.7374

Tabel 8. Pengelompokan data ke-2

NO	C1	C2
1	1	0
2	1	0
3	1	0
4	1	0
5	1	0
6	1	0
7	1	0
8	0	1
9	1	0
10	0	1
11	1	0
12	1	0
13	1	0
14	1	0
15	1	0
16	1	0
17	1	0
18	1	0
19	1	0
20	1	0

Iterasi dengan cara yang sama akan dilakukan sampai tidak ada perubahan data yang terpada pengelompokan *cluster* seperti pada tabel 4.3. dalam penelitian ini dikarenakan banyaknya data maka iterasi dilakukan sampai 9 kali. Pada iterasi ke-9 diperoleh *cluster* yang tidak berubah dari *cluster* iterasi ke -8 dimana nilai C1 = 657 dan C2 = 70 sama dengan nilai pada iterasi ke-9. Hasilnya seperti pada tabel di bawah.

Tabel 9. Tabel Iterasi dan pengelompokan data ke-9

No	nama barang	Fe	Ma	AP	C1	C2	arak terpek dek
----	-------------	----	----	----	----	----	-----------------

1	0029 alas panas motif	0	0	4	7.5	67	7.492405
2	3273 alas panas motif	0	0	1	8.3	68	8.251972
3	393#15*15 alas panas persegi	0	0	1	8.3	68	8.251972
4	539 alas panas circle	0	0	1	8.3	68	8.251972
5	alas kaki	0	0	30	27	64	26.57992
6	2's qiangshun perekat tikus	0	0	0	8.7	69	8.722462
7	items fresh non barcode	1	0	2	7.2	67	7.229278
8	02 shunfeng baskom	58	31	29	64	13	13.23503
9	057 centa t.sendok	7	5	9	4.8	58	4.776981
10	6884-baskom	36	16	17	35	26	26.43202
11	802 ke wei baskom	7	1	2	5.1	61	5.088477
12	8032 baskom	11	7	11	8.8	53	8.767975
13	3s eveready s91 bp20 sil 3	3	1	0	6.6	66	6.576495
14	3s eveready s92 bp20 sil 4	1	0	0	8.1	68	8.138493
15	4's eveready 1012 bp20 aaa mrh 1	16	9	0	12	52	12.12058
16	4's eveready 1015 bp20 aa mrh kcl 2	30	11	0	26	40	25.66084
17	energizer e91 aa bp2+1 max 1	16	20	0	19	51	18.78021
18	energizer e92 aaa bp2+1max 2	16	7	0	12	53	11.64025
19	1199 bola sport	0	0	6	7.6	67	7.626502
20	7788 bola smiley	0	0	2	7.9	68	7.88091

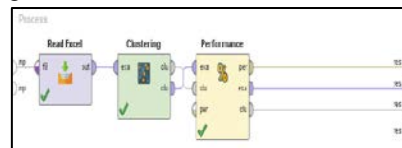
Tabel 10. Kelompok data ke-9

NO	C1	C2
1	1	0
2	1	0
3	1	0
4	1	0
5	1	0
6	1	0

7	1	0
8	0	1
9	1	0
10	0	1
11	1	0
12	1	0
13	1	0
14	1	0
15	1	0
16	1	0
17	1	0
18	1	0
19	1	0
20	1	0

Implementasi dengan Rapidminer

Pengolahan data menggunakan rapidminer seperti ditampilkan pada gambar di bawah:



Gambar 4. Pemodelan K-means pada Rapidminer

Pemodelan *cluster* ing dengan algoritma k-means dengan jumlah *cluster* sebanyak 2 buah maka dengan menggunakan rapidminer *cluster* yang terbentuk juga sebanyak 2, dengan *cluster* 0 sebanyak 657 produk, *cluster* 1 sebanyak 70 produk dari jumlah produk 727 item.

Cluster Model

Cluster 0: 657 items
 Cluster 1: 70 items
 Total number of items: 727

Gambar 5. Hasil *cluster* K-means menggunakan Rapidminer

Attribute	cluster_0	cluster_1
Fe	5.423	61.243
Ma	5.146	19.100
AP	4.493	24.200

Gambar 6 Hasil perhitungan jarak *cluster* dengan centroid

Pembahasan

Clustering dengan algoritma k-means yang dilakukan terhadap data transaksi penjualan PT Batamas Niaga Jaya dengan cara manual dilakukan sebanyak 9 kali iterasi yang terdiri dari 2 *cluster* yaitu C0 dan C1, C0 merupakan produk yang tidak laris dan C1 merupakan produk yang laris, pengelompokan ini dapat dilihat pada nilai centroid iterasi ke-8 dan ke-9 dimana nilai centroid ini diperoleh dari nilai rata-rata penjualan untuk *cluster* 0 february, maret dan april jauh di bawah nilai centroid *cluster* 1, seperti pada tabel 4.6 di bawah.

Tabel 11. Nilai Centroid Ke-8 dan Ke-9

Titik Pusat	Februari	Maret	April
Cluster 0	5.423135464	5.146118721	4.493150685
Cluster 1	61.24286	19.1	24.2

Proses *clustering* untuk 727 produk yang dianalisis dengan algoritma K-means memperoleh hasil C0 sebanyak 657 item dan C1 sebanyak 70 item yang diuraikan sebagai berikut:

1. *Cluster* C0 sebanyak 657 dengan no produk: 1-7, 9, 11-20, 26-193, 195,197-209, 211-225,227-230, 232-267, 271, 272, 278, 279, 286, 289, 293, 294, 296, 304, 305, 308,

309, 310, 312, 313, 318-321, 323-331, 333-338, 341, 343, 345-368, 370-384, 386-397, 399-407, 409-432, 434, 436-480, 482-488, 490-496, 498, 499, 501,502, 504-522, 524-587, 589, 591-635, 637, 639-661, 663-715, 717,718, 720-727.

2. *Cluster* C1 sebanyak 70 produk dengan nomor produk: 8, 10, 25,31,194, 196, 210, 226,231, 268, 269, 270, 273-277, 280-285, 287, 288, 290, 291, 292, 295, 297-303, 306, 307, 311, 314-317, 322, 332,339, 340, 342, 344, 369, 385, 398, 408, 433, 435, 481, 489, 497, 500, 503, 523, 588, 590, 636, 638, 662, 716, 719,.

Clustering yang dilakukan dengan implementasi rapidminer memperoleh hasil seperti yang ditampilkan pada gambar 4.2 dimana jumlah *cluster* sebanyak 2 yaitu *cluster* 0 dan *cluster* 1, dan disebutkan juga bahwa *cluster* 0 sebanyak 657 *items* dan *clusters* 1 sebanyak 70 *items*. Nilai centroid yang dihasilkan juga dengan menggunakan rapidminer juga sama dengan hasil yang diperoleh dengan cara manual. Hal ini memperkuat tingkat akurasi dari pengelompokan produk yang tidak laris dan produk yang laris pada PT Batamas Niaga Jaya.

Dengan terkelompoknya produk tidak laris dan produk laris pada PT Batamas Niaga Jaya khususnya untuk Toko Serba 10.000 dan Toko Serba 7.500 maka diharapkan dapat mengurangi kesulitan yang dialami diantaranya kekurangan stok produk yang laris dan menumpuknya produk-produk yang tidak laris. Produk-produk yang berada pada *cluster* 1 jumlah stoknya dapat diperbanyak sedangkan untuk produk-produk dengan *cluster* 0 jumlah stoknya dapat dikurangi, selain itu pihak

manajemen PT Batamas Niaga Jaya juga harus melakukan berbagai strategi untuk meningkatkan penjualan mengingat bahwa produk yang tidak laris jauh lebih banyak dari produk yang laris oleh karena itu harus ditempuh strategi khusus untuk mengatasi masalah penjualan jenis produk pada cluster 0.

V. Simpulan

Berdasarkan hasil pengklasifikasian dengan algoritma K-Means maka dapat diberikan beberapa kesimpulan sebagai berikut:

1. Hasil *clustering* K-means yang dilakukan memperoleh 2 jumlah *cluster*
2. Pengelompokan produk berdasarkan data penjualan pada PT Batamas Niaga Jaya dengan k-means memperoleh 2 kelompok produk yaitu produk yang tidak laris dengan kelompok C0 dan kelompok produk yang laris dengan kelompok C1.
3. Kelompok C0 produk yang tidak laris terdiri dari 657 produk dan kelompok C1 terdiri dari 70 produk yang merupakan produk yang laris.

Ucapan Terimakasih

Dengan terselesaikannya karya ilmiah ini, penulis mengucapkan terimakasih yang sedalam-dalamnya kepada semua pihak yang sudah membantu dalam penyelesaian penelitian ini baik dalam bentuk moril maupun materil.

Daftar Pustaka

- [1] I. Sumadikarta and E. Abeiza, "PENERAPAN ALGORITMA K-MEANS PADA DATA MINING UNTUK MEMILIH PRODUK DAN PELANGGAN POTENSIAL (Studi Kasus : PT Mega Arvia Utama)," *Satya Inform.*, vol. 1, No.1, p. 14, 2014.
- [2] A. F. Manampiring and I. Trang, "Pengaruh Produk , Harga , Promosi Dan Tempat Terhadap Keputusan," *EMBA*, vol. 4, No.1, no. 1, pp. 472–483, 2016.
- [3] A. A. Ulus, "Bauran Pemasaran Pengaruhnya Terhadap Keputusan Pembelian Mobil Daihatsu PT Astra Internasional Manado," *J. EMBA*, vol. 1, no. 4, pp. 1134–1144, 2013.
- [4] R. Oktavianus, "ANALISIS DAN IMPLEMENTASI SISTEM INFORMASI PENJUALAN PADA PD . CITRA KOMPUTER PONTIANAK," vol. 1, no. 1, pp. 28–29, 2011.
- [5] T. Ariyanti and S. R. Fuadati, "Pengaruh Strategi Promosi Terhadap Penjualan Produk Di Pt . United Indo Surabaya," *J. Ilmu Ris. Manaj.*, vol. 3. No.10, no. 10, pp. 1–19, 2014.
- [6] B. M. Metisen and H. L. Sari, "ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA," vol. 11, no. 2, pp. 110–118, 2015.
- [7] A. Bahar, B. Pramono, and L. H. S. Sagala, *Penentuan strategi penjualan alat-alat tattoo di studio sonyxtattoo menggunakan metode*, vol. 2, no. 2. 2016.
- [8] M. Neel B, "Predictive Data mining and discovering hidden values of Data warehouse," *ARPN J. Syst. Softw.*, vol. 1, no. 1, pp. 1–5, 2011.
- [9] F. Chen, P. Deng, J. Wan, D. Zhang, A. V Vasilokas, and X. Rong, "Data mining for the internet of things: literature review and challenges," *Int. J. Distrib. Sens. Networks*, vol. 2015, no. i, 2015.
- [10] S. N. Kohail and A. M. El-halees, "Implementation of Data Mining

- Techniques for Meteorological Data Analysis (A case study for Gaza Strip),” *Int. J. Inf. Commun. Technol. Res.*, vol. 1, no. 3, pp. 96–100, 2011.
- [11] Mardiani, “Perbandingan Algoritma K-Means dan EM untuk Clusterisasi Nilai Mahasiswa Berdasarkan Asal Sekolah,” *Citec J.*, vol. 1, No.4, pp. 317–318, 2008.
- [12] F. Shabir and A. Rahman M, “REKOMENDASI PEMBELIAN PERSONAL KOMPUTER DENGAN METODE RANKED CLUSTERING,” *J. Ilm. Ilk.*, vol. 8, no. Agustus, pp. 119–125, 2016.
- [13] Sismadi, “Vol. IX No.1, Maret 2013 Techno Nusa Mandiri PERBANDINGAN TINGKAT AKURASI PEMETAAN ALGORITMA K-MEANS DAN FUZZY C-MEANS NASABAH ASURANSI Jiwa SYARIAH,” *Techno Nusa Mandiri*, vol. IX, no. 1, pp. 109–119, 2013.
- [14] B. R. C. T. I, A. A. Gafar, N. Fajriani, U. Ramdani, F. R. Uyun, Y. P. P, and N. Ransi, “Implementasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan,” *Semin. Nas. Ris. Kuantitatif Terap.*, no. April, pp. 58–62, 2017.