

Algoritma Asosiasi *K-Means* dan *FP-Growth* untuk Analisis Keranjang Pasar pada Penjualan Produk Alumunium

Ela Nurelasari^{1,*}

¹ Manajemen Informatika; AMIK BSI Jakarta; Jl. R.S. Fatmawati No.24, Pondok Labu, Jakarta Selatan, telp/fax (021)7500282/(021)7513790; e-mail: ela.eur@bsi.ac.id

* Korespondensi: e-mail: ela.eur@bsi.ac.id

Diterima: 22 April 2017; Review: 29 April 2017; Disetujui: 06 Mei 2017

Cara sitasi: Nurelasari E. 2017. Algoritma Asosiasi *K-Means* dan *FP-Growth* untuk Analisis Keranjang Pasar pada Penjualan Produk Alumunium. *Information Management For Educators And Professionals*. 1 (2): 179 – 186.

Abstrak: Alumunium banyak digunakan di berbagai bidang, sehingga bermunculan perusahaan-perusahaan untuk mengolah aluminium. Semakin banyak usaha-usaha dalam bidang yang sama menimbulkan persaingan usaha. Untuk mengatasinya diperlukan strategi pemasaran yang baik. Salah satu penelitian yang banyak dilakukan yaitu menganalisis keranjang belanja untuk menentukan rekomendasi produk kepada pelanggan. Banyak peneliti menggunakan aturan asosiasi dengan apriori atau *fp-growth* dalam menganalisis keranjang belanja. Tetapi untuk dataset yang besar, hasil dari penerapan asosiasi menjadi kurang akurat. Oleh karena itu dataset yang besar akan disegmentasi dengan *k-means* agar dataset yang besar dibagi menjadi beberapa segmen yang lebih kecil. Hasil yang didapat dalam menganalisis keranjang belanja dengan menerapkan algoritma *k-means* dan algoritma *fp-growth* terbukti dapat meningkatkan akurasi dari 70% menjadi 90 %,80% dan 90%. Rekomendasi produk yang tepat dapat membantu dalam strategi pemasaran, khususnya dalam bidang promosi produk dan untuk membantu perencanaan produksi produk.

Kata kunci: algoritma *k-means*, algoritma *fp-growth*, keranjang belanja, rekomendasi produk.

Abstract: *Aluminum is widely used in a variety of things, so that the emerging companies to process aluminum. More and more businesses in the same pose competition. To overcome this need a good marketing strategy. One study done analyzing Market Basket to determine product recommendations to customers. Many researchers associate a priori approach or fp-growth in analyzing market basket. But for a large dataset, the results of the application of the association becomes less accurate. Therefore, a large dataset to be segmented k-means that large datasets are divided into several smaller segments. Analyzing the results obtained in the market basket by applying k-means algorithm and fp-growth algorithm is shown to improve the accuracy of 70% to 90%, 80% and 90%. Appropriate product recommendations to assist in the marketing strategy, especially in the field of promotional products and to assist production planning.*

Keywords: *k-means algorithm, fp-growth algorithm, market basket, product recommendation.*

1. Pendahuluan

Aluminium merupakan konduktor listrik dan panas yang baik, kuat, dapat ditempa menjadi lembaran, kawat, batangan dan tahan korosi, sehingga banyak digunakan untuk berbagai bidang, yaitu untuk mengolah air, membuat kertas, farmasi, sebagai bahan di dalam pesawat terbang, konstruksi, mobil dan mesin, kemasan makanan dan peralatan masak, dan lain-lain (EFSA, 2008, hal. 12). Oleh karena itu mulai bermunculan perusahaan-perusahaan untuk mengolah aluminium tersebut menjadi bermacam-macam produk, salah satunya perusahaan manufaktur aluminium di daerah Tangerang yang memproduksi berbagai macam produk aluminium untuk peralatan rumah tangga, yaitu seperti berbagai macam kunci untuk pintu, tangki, kitchen set, peralatan makan, tangga, jemuran, pintu kamar mandi, meja, lemari, batangan-batangan aluminium untuk bahan bangunan, dan lain-lain. Segmen pasarnya B to B (*Bussines to Bussines*) yaitu distributor seperti Mitra 10 dan Dunia Bangunan, modern market seperti *Hypermart*, *Carefour*, dan *Lotte Mart*, dan juga untuk toko-toko bangunan lainnya untuk dalam negeri maupun *export*. Dengan segmen pasar yang luas sehingga persaingan usaha pun menjadi lebih luas. Untuk mengatasi persaingan usaha maka diperlukan strategi pemasaran yang tepat (Liao & Triantaphyllou, 2007, hal. 5). Untuk meningkatkan ketepatan dalam strategi pemasaran telah banyak penelitian dilakukan dalam bidang penjualan, salah satunya yaitu analisis keranjang pasar. Munculnya penjualan elektronik semakin meningkatkan kemajuan dalam berbagai teknik dan model yang dimaksudkan untuk meningkatkan kepuasan pelanggan. Salah satunya mencoba untuk menentukan keinginan Pelanggan (Cavique, 2007, p. 1). Hal ini cukup penting karena dapat menjaga relasi dengan pelanggan (Yang & Padmanabhan, 2003, hal. 1), dapat membantu rekomendasi produk dan promosi produk sehingga strategi pemasaran pun menjadi lebih tepat.

Banyak penelitian menganalisis keranjang pasar untuk menentukan pola asosiasi menggunakan algoritma Apriori (Liao & Triantaphyllou, 2007), seperti Lee and Lee mencari pola penjualan dan Wong *et al.* mencari pola perubahan kebiasaan pelanggan, dan lain-lain. Selain itu banyak juga peneliti-peneliti yang menggunakan algoritma lain, seperti Christidis *et al.* melakukan penelitian pada e-commerce dengan menggunakan latent topic models pada data transaksi histori pasar pelanggan sehingga dapat membantu menentukan rekomendasi produk kepada pelanggan menjadi lebih efektif bahkan pada dataset dan itemset yang besar, Cavique untuk menemukan pola penjualan data sekuensial dalam data yang besar dengan waktu cukup singkat menggunakan metode ramex. Untuk menganalisis keranjang pasar, pendekatan yang biasa digunakan adalah aturan asosiasi. Tetapi ada sejumlah masalah teknis yang berhubungan dengan teknik rekomendasi yang paling umum. Aturan asosiasi cenderung mengabaikan itemset besar, dan rekomendasi item kurang tepat karena informasi tentang produk ritel tidak tersedia (Christidis, Apostolou, & Mentzas, 2010), sehingga untuk data yang besar hasilnya menjadi kurang akurat. Dataset yang sudah dibuat branch atau diklaster dapat membantu mempermudah proses mencari rekomendasi produk (Cavique, 2007), dikarenakan jumlah data dibagi menjadi beberapa bagian.

Pada penelitian ini, akan dilakukan segmentasi dengan menggunakan algoritma *K-means* pada data penjualan, kemudian setiap segmen akan dilakukan pendekatan asosiasi menggunakan algoritma *Fp-growth*, dengan demikian dataset yang akan diasosiasi menjadi lebih kecil, sehingga rekomendasi produk kepada pelanggan menjadi lebih akurat.

2. Metode Penelitian

Metode penelitian yang dilakukan adalah metode penelitian eksperimen, dengan tahapan penelitian seperti berikut:

1. Pengumpulan Data

Pertama-tama dilakukan pengumpulan data-data yang berhubungan dengan masalah penelitian. Jenis pengumpulan data dibagi menjadi 2, yaitu: data sekunder adalah data yang bersumber dari hasil penelitian orang lain yang dibuat untuk tujuan berbeda tapi bisa dimanfaatkan dan data Primer adalah data yang dikumpulkan oleh pene iti langsung dari sumber utamanya atau data hasil eksperimen yang kita lakukan.

Data penelitian ini diperoleh dari data transaksi penjualan aluminium pada sebuah perusahaan manufaktur aluminium di daerah tangerang yang memproduksi berbagai macam produk aluminium untuk peralatan rumah tangga, yaitu seperti berbagai macam kunci untuk pintu, tangki, kitchen set, peralatan makan, tangga, jemuran, pintu kamar

mandi, meja, lemari, batangan-batangan alumunium untuk bahan bangunan, dan lain-lain. Segmen pasarnya B to B (*Bussines to Bussines*) yaitu distributor seperti Mitra 10 dan Dunia Bangunan, modern market seperti *Hypermart*, *Carefour*, dan *Lotte Mart*, dan juga untuk toko-toko bangunan lainnya untuk dalam negeri maupun *export*. Data yang digunakan adalah data penjualan mulai dari tanggal 1 januari 2006 sampai 31 Desember 2011 dengan atribut-atribut yaitu: kode penjualan, tanggal, no CO, no surat jalan, kode pelanggan, kode produk, qty, harga, diskon, diskon2, ppn, jatuh tempo, total jual, total bayar, tanggal pelunasan, kode pelunasan.

2. Pengolahan Awal Data

Data yang sudah dikumpulkan kemudian diolah dengan cara dilakukan seleksi atribut, menghapus data yang tidak lengkap dan data yang *redundant* agar mengurangi data yang tidak relevan.

a. Data Preprocessing

Data Preprocessing yaitu dilakukan pembersihan dan persiapan data untuk menghilangkan konsistensi data, data tidak lengkap dan redundant data yang terdapat pada data awal.

b. Seleksi Atribut

Seleksi atribut dilakukan untuk memilih atribut mana saja yang di butuhkan dalam proses menganalisis keranjang pasar, sehingga data yang diambil menjadi semakin sedikit. Atribut yang diambil dari data penjualan yaitu: *custcode*, *itemcode* dan *qty*.

Tabel 1. Data setelah diseleksi atribut

custcode	itemcode	Qty
22070	TG310031	3
22069	TG310041	2
22188	TG310011	4
22188	TG310031	1
22188	TG310051	4
22188	TG310061	1
22188	TG310061	5
22236	TG310071	2
22700	TG310031	5
22700	TG310041	3
22036	TG310031	10
22068	TG310011	4
22068	TG310031	4
22068	TG310032	1
22068	TG310071	1
22036	TG310031	2
22379	TG310011	1
22379	TG310012	1
22379	TG310041	1
22188	TG310011	2
22188	TG310012	1
22188	TG310031	10
22188	TG310051	1
22188	TG310061	4
22236	TG310031	5
22236	TG310031	4
22639	TG310012	3
22639	TG310012	1
22639	TG310031	3

Sumber : Hasil Penelitian (2017)

c. Pembuatan Dataset

Untuk memudahkan proses segmentasi data penjualan berdasarkan total penjualan yang kemudian hasil segmentasi akan dilakukan proses asosiasi, maka data-data yang sudah diseleksi akan dibuat pivot tabel, sehingga atributnya menjadi: custcode, qty produk1, qty produk2, qty produk3, qty produk n, seperti terlihat pada tabel 2.

Tabel 2. Dataset

custcode	sum(qty)_DL150	sum(qty)_DL200	sum(qty)_DL175	sum(qty)_A160S	sum(qty)_A180S
12849	?	50	?	40	240
12853	5	?	?	?	?
12857	29	75	22	3	4
12858	176	126	92	31	33
12859	76	103	76	19	29
12860	20	5	10	?	?
12861	10	20	15	5	5
12877	42	47	31	?	5
12881	59	99	196	8	20
12883	2	10	?	?	3
12886	21	5	18	5	5

Sumber : Hasil Penelitian (2017)

3. Metode Yang Diusulkan

Metode yang diusulkan pada penelitian ini mulai dari pengumpulan data awal sampai dengan rekomendasi produk per segmen 4.

4. Eksperimen dan Pengujian Metode

Eksperimen dan pengujian metode dilakukan menggunakan Rapid Miner 5.

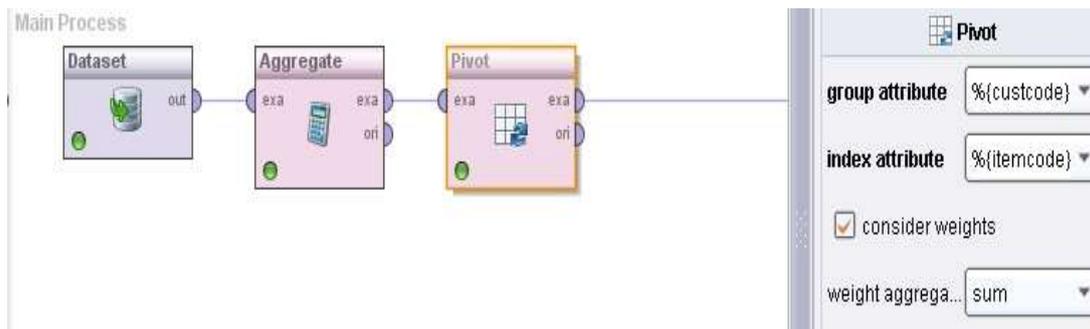
a. Segmentasi data Penjualan

Dataset diambil dari data penjualan dalam bentuk excel diimport ke dalam rapid miner diimport seperti terlihat pada gambar 1. Kemudian data yang sudah diimport dibuat menjadi pivot untuk memudahkan dalam proses asosiasi

batch	trndate	no_sjln	custcode	itemcode	qty	harsat
100100000E	Mon Jan 04	912017098	22482	KT334037	1	1541760
100100000E	Mon Jan 04	912017097	22482	KT310018	6	1593715
100100000E	Mon Jan 04	912017097	22482	KT320019	5	1147475
1001000007	Mon Jan 04	912017091	22482	KT310006	6	395738
1001000007	Mon Jan 04	912017091	22482	KT310019	4	883099
1001000007	Mon Jan 04	912017091	22482	KT320001	3	745942
1001000007	Mon Jan 04	912017091	22482	KT320026	3	1268742
1001000007	Mon Jan 04	912017091	22482	KT351011	8	471905

Sumber : Hasil Penelitian (2017)

Gambar 1. Import Data



Sumber : Hasil Penelitian (2017)

Gambar 2. Proses Pivot Data

Role	Name	Type	Statistics	Range
regular	custcode	nominal	mode = 12011 (1), lea:	12011 (1), 12059 (1), 1
regular	sum(qty)_DL150	numeric	avg = 91.178 +/- 176.8:	[2.000 ; 975.000]
regular	sum(qty)_DL200	numeric	avg = 81.472 +/- 162.1:	[2.000 ; 982.000]
regular	sum(qty)_DL175	numeric	avg = 66.565 +/- 108.2:	[1.000 ; 611.000]
regular	sum(qty)_A160S	numeric	avg = 51.417 +/- 114.7:	[1.000 ; 503.000]
regular	sum(qty)_A180S	numeric	avg = 94.067 +/- 212.7:	[1.000 ; 1003.000]
regular	sum(qty)_P180RABF	numeric	avg = 15.143 +/- 26.36:	[1.000 ; 74.000]
regular	sum(qty)_P180RAGR	numeric	avg = 30 +/- 42.360	[3.000 ; 104.000]
regular	sum(qty)_P180RABC	numeric	avg = 20.167 +/- 25.35:	[3.000 ; 63.000]
regular	sum(qty)_A085B	numeric	avg = 46.273 +/- 74.35:	[2.000 ; 240.000]
regular	sum(qty)_A085S	numeric	avg = 58.211 +/- 118.5:	[1.000 ; 406.000]

Sumber : Hasil Penelitian (2017)

Gambar 3. Hasil Pivot Data

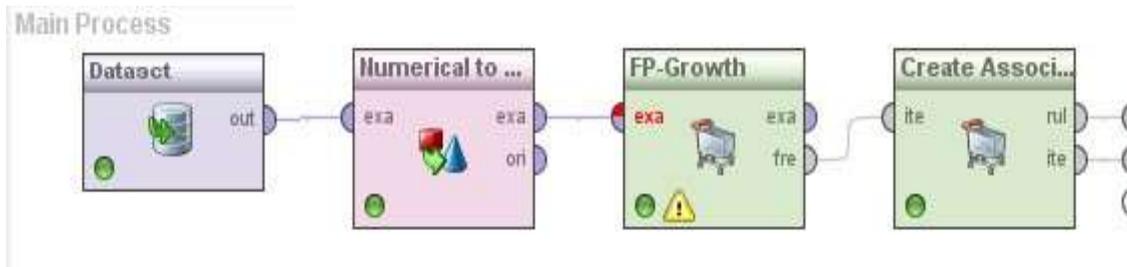
Attribute	cluster_0	cluster_1	cluster_2
FG310011	4	0	0
FG310012	0	0	0
FG310031	1	10	2
FG310032	0	0	0
FG310041	0	0	0
FG310051	4	0	0
FG310061	6	0	0

Sumber : Hasil Penelitian (2017)

Gambar 4. Hasil Segmentasi menggunakan K-means

b. Rekomendasi Produk

Setiap segmen akan dilakukan asosiasi dengan *Fp-growth* untuk menghasilkan rekomendasi produk. Langkah-langkahnya yaitu dataset masing- masing segmen akan diproses numerical to binominal tujuannya untuk mengubah nilai numerik menjadi binominal (true dan false), karena *Fp-growth* hanya dapat memproses atribut-atribut binominal. Setelah proses *Fp-growth* untuk menghasilkan rekomendasi produk maka dilanjutkan dengan proses create association rule.



Sumber : Hasil Penelitian (2017)

Gambar 5. Hasil Asosiasi menggunakan Fp-growth

5. Evaluasi dan Validasi Hasil

Evaluasi hasil yang didapat setelah dilakukan eksperimen dan pengujian metode yaitu dengan mengukur akurasi dari hasil algoritma yang digunakan.

3. Hasil dan Pembahasan

Hasil penelitian ini menggunakan *precision*, *recall* dan F1 untuk mengukur tingkat akurasi.

3.1. Segmentasi Data Penjualan

Data yang sudah ada akan disegmentasi dengan algoritma *k-means*, dengan atribut yang akan digunakan yaitu *custcode*, *qtyproduk1*, *qtyproduk2*, ..., *qtyproduk n* dapat dilihat di tabel 3.

Tabel 3. Contoh data untuk segmentasi

No	Produk 1	Produk 2	Produk 3
1	0	3	0
2	0	0	2
3	4	1	0
4	0	0	1
5	0	5	0
6	0	10	0
7	4	4	0
8	0	2	0
9	1	0	1
10	2	9	0

Sumber : Hasil Penelitian (2017)

Secara acak akan dipilih data sejumlah dengan nilai *k* adalah 3. Contoh data yang dipilih adalah nomor 1, 3, dan 5 sebagai pusat segmentasi pertama. Nilai *euclidean* diperhitungkan untuk setiap data ke dalam ketiga titik tersebut.

Setelah semua titik *euclidean* diperhitungkan, maka kluster pertama yang berpusat pada data ke-1 terdiri dari data 1, 2, 4, 8, dan 9. Kluster kedua berpusat pada ke-3 terdiri data 3 dan 7.

Klaster ketiga berpusat pada data ke-5 terdiri dari data 5, 6, dan 10. Hasil perhitungan titik *euclidean* dapat dilihat pada Tabel 4

Tabel 4. Hasil perhitungan titik euclidean pertama

No	1	3	5
1	0,00	4,47	2,00
2	3,61	4,58	5,39
3	4,47	0,00	5,66
4	3,16	4,24	5,10
5	2,00	5,66	0,00
6	7,00	9,85	5,00
7	4,12	3,00	4,12
8	1,00	4,12	3,00
9	2,24	4,12	4,12
10	6,32	8,25	4,47

Sumber : Hasil Penelitian (2017)

3.2. Implikasi Penelitian

Pendekatan asosiasi merupakan pendekatan yang banyak digunakan untuk menganalisis keranjang pasar. Tapi untuk data yang besar, pendekatan asosiasi menjadi kurang akurat, sehingga dilakukan segmentasi agar dataset menjadi lebih kecil, dan hasil akurasi yang didapat juga menjadi lebih meningkat. Berikut adalah perbandingan dari hasil *Fp-growth + K-means* dan *Fp-growth* saja dapat dilihat pada tabel 22.

Tabel 5. Perbandingan akurasi model *Fp-growth* pada semua segmen dan tanpa segmentasi

K1	K2	K3	Tanpa segmen
75%	67%	95%	75%
78%	71%	71%	71%
75%	75%	80%	71%
89%	82%	78%	78%
62%	62%	62%	50%
92%	83%	92%	73%
77%	83%	83%	83%
71%	75%	90%	89%
77%	77%	77%	40%
78%	82%	82%	57%

Sumber : Hasil Penelitian (2017)

Hasil dari penelitian ini ditemukan perbedaan rekomendasi produk pada ketiga segmen dan juga pada dataset yang tidak dilakukan segmentasi, yaitu:

- a. Tingkat akurasi tertinggi pada segmen pertama yaitu mencapai 92% dan tingkat akurasi terendah mencapai 62%.
- b. Tingkat akurasi tertinggi pada segmen kedua yaitu mencapai 83% dan tingkat akurasi terendah mencapai 62%.

- c. Tingkat akurasi tertinggi pada segmen ketiga yaitu mencapai 95% dan tingkat akurasi terendah mencapai 62%.
- d. Tingkat akurasi tertinggi pada dataset tanpa segmentasi yaitu mencapai 89% dan tingkat akurasi terendah mencapai 40%.

4. Kesimpulan

Dari penelitian yang dilakukan, segmentasi dilakukan sebelum penerapan aturan asosiasi untuk menemukan produk yang sering muncul untuk meningkatkan akurasi rekomendasi produk kepada pelanggan. Segmentasi dilakukan menggunakan algoritma *K-means* untuk membagi dataset yang besar ke dalam beberapa dataset yang lebih kecil. Peningkatan akurasi rekomendasi produk dapat dilihat dari peningkatan akurasi yang diukur dengan menggunakan *F1*, *precision* dan *recall*, untuk model algoritma *Fp-growth + K-Means* pada segmen pertama akurasi tertinggi mencapai 92% dan akurasi terendah 62%, pada segmen kedua akurasi tertinggi mencapai 83% dan akurasi terendah 62%, pada segmen ketiga akurasi tertinggi mencapai 95% dan akurasi terendah 62%, dan untuk model algoritma *Fp-growth* tanpa disegmentasi dahulu akurasi tertinggi mencapai 89% dan akurasi terendah mencapai 40%.

Referensi

- Cavique, L. (2007). A Network Algorithm to Discover Sequential Patterns. *ESCS- IPL, Portugal*
- Christidis K, Apostolou D, Mentzas G. 2010. Exploring Customer Preferences with Probabilistic Topics Models. *National Technical Univesity of Athens* .
- EFSA. 2008. Scientific Opinion of the Panel on Food Additives, Flavourings, Processing.
- Han J, Kamber M. 2007. *Data Mining: Concepts and Techniques*. San Fransisco: Elsevier Inc.
- Hendriani L. 2007. Inovasi Program Promosi untuk Sales. *Jakarta: AcNielsen* .
- Larose DT. 2005. *Discovering Knowledge in Data an Introduction to Data Mining*. New Jersey: John Wiley & Sons, Inc., Hoboken.
- Li Y, Meiyun Z, Yang B. 2005. Analysis and Design of E-Supermarket Shopping . *ACM 1-59593-112-0/05/08* .
- Liao TW, Triantaphyllou E. 2007. *Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications* Vol. 6 (5) Toh Tuck Link: World Scientific Publishing Co. Pte. Ltd.
- Myatt GJ. 2007. *Making Sense of Data : A Practival Guide to Exploratory Data Analysis and Data Mining*. Hoboken: John Wiley & Sons.
- Witten IH. Eibe F, Hall MA. 2011. *Data Mining: Practical Machine Learning Tools and Techniques 3D Edition*. United State.
- Wu X, Kumar V. 2009. *The Top Ten Algorithms in Data Mining*.
- Taylor & Francis Group LLC, Xu R, Wunsch II DC. 2009. *Clustering*. Hoboken: John Wiley & Sons, Inc.
- Yang Y, Padmanabhan B. 2003. A Hierarchical Pattern Based Clustering Algorithm for. *IEEE International Conference on Data Mining*.