

Artikel Teknis

Penambangan data untuk industri ritel online: Studi kasus Segmentasi pelanggan berbasis model RFM menggunakan penambangan data

Diterima (dalam bentuk revisi): 18 Juli 2012

Daqing Chen adalah

dosen senior di Departemen Informatika, Fakultas Bisnis, Universitas London South Bank, London, Inggris. Dia terutama memberikan kuliah tentang penambangan data dan intelijen bisnis pada program BSc dan MSc. Minat penelitiannya meliputi penambangan data, pemasaran berbasis data, dan intelijen bisnis yang berpusat pada pelanggan. Dalam beberapa tahun terakhir, ia telah terlibat dalam beberapa proyek penambangan data berorientasi bisnis di berbagai sektor bisnis.

Sai Laing Sain saat ini

adalah mahasiswa BSc di Departemen Informatika, Fakultas Bisnis, London South Bank University, London, Inggris.

Kun Guo

saat ini adalah mahasiswa PhD di Departemen Teknik Sipil, Fakultas Teknik, Sains, dan Lingkungan Binaan di London South Bank University, London, Inggris. Minat akademisnya meliputi pemodelan numerik, algoritma kecerdasan buatan, dan penambangan data.

ABSTRAK Banyak pengecer daring kecil dan pendatang baru di sektor ritel daring sangat ingin mempraktikkan penambangan data dan pemasaran yang berpusat pada konsumen dalam bisnis mereka, tetapi secara teknis tidak memiliki pengetahuan dan keahlian yang diperlukan untuk melakukannya. Dalam artikel ini, disajikan studi kasus penggunaan teknik penambangan data dalam intelijen bisnis yang berpusat pada pelanggan untuk pengecer daring. Tujuan utama analisis ini adalah untuk membantu bisnis lebih memahami pelanggannya dan karenanya menjalankan pemasaran yang berpusat pada pelanggan secara lebih efektif. Atas dasar model Keterkinian, Frekuensi, dan Moneter, pelanggan bisnis telah disegmentasikan ke dalam berbagai kelompok yang bermakna menggunakan algoritma pengelompokan k-means dan induksi pohon keputusan, dan karakteristik utama konsumen di setiap segmen telah diidentifikasi dengan jelas. Oleh karena itu, serangkaian rekomendasi selanjutnya diberikan kepada bisnis tentang pemasaran yang berpusat pada konsumen.

SAS Enterprise Guide dan SAS Enterprise Miner digunakan dalam studi ini.

Jurnal Pemasaran Basis Data & Manajemen Strategi Pelanggan (2012) **19**, 197 – 208.

doi: 10.1057/dbm.2012.17 ; dipublikasikan online 27 Agustus 2012

Kata kunci: ritel online, pemasaran yang berpusat pada pelanggan, penambangan data, segmentasi pelanggan, model RFM, pengelompokan k-means

Korespondensi: Daqing
Chen Departemen
Informatika, Fakultas Bisnis,
London South Bank University,
London, Inggris Email:
chend@lsbu.ac.uk

PENGANTAR Selama

10 tahun terakhir, kita telah menyaksikan peningkatan penjualan ritel online yang stabil dan kuat. Menurut Interactive Media

Menurut data dari Retail Group (IMRG), pembelian online di Inggris menghabiskan sekitar £50 miliar pada tahun 2011, meningkat lebih dari 5000 persen dibandingkan tahun 2000.



tahun 2000.¹ Peningkatan yang luar biasa ini

penjualan menunjukkan bahwa cara konsumen berbelanja dan menggunakan layanan keuangan telah berubah secara mendasar.

Dibandingkan dengan belanja tradisional di toko ritel, belanja online memiliki beberapa karakteristik unik: setiap proses dan aktivitas belanja pelanggan dapat dilacak secara instan dan akurat, setiap pelanggan

Pesanan biasanya dikaitkan dengan alamat pengiriman dan alamat penagihan, dan setiap pelanggan memiliki akun toko online dengan kontak penting dan

informasi pembayaran. Karakteristik belanja online yang diinginkan dan khusus ini telah memungkinkan pengecer online untuk memperlakukan setiap pelanggan sebagai individu dengan pemahaman yang dipersonalisasi terhadap setiap pelanggan dan membangun kecerdasan bisnis yang berpusat pada pelanggan.

Sehubungan dengan bisnis yang berpusat pada pelanggan intelijen, pengecer online biasanya memperhatikan masalah bisnis umum berikut:

- Halaman web produk/item apa saja yang pernah dikunjungi pelanggan? Berapa lama pelanggan berada di setiap halaman web, dan dalam urutan apa pelanggan mengunjungi sekumpulan halaman web produk?
- Siapa pelanggan yang paling/paling tidak berharga? untuk bisnis? Apa saja perbedaannya? karakteristik mereka?
- Siapa saja pelanggan yang paling/paling tidak setia, dan bagaimana karakteristik mereka?
- Apa saja pola perilaku pembelian pelanggan? Produk/barang apa saja yang sering dibeli pelanggan secara bersamaan? Dalam urutan apa produk tersebut dibeli?
- Jenis pelanggan mana yang lebih mungkin menanggapi surat promosi tertentu? dan
- Bagaimana pola penjualan dalam berbagai perspektif seperti produk / item, wilayah, dan waktu (mingguan, bulanan, triwulanan, tahunan, dan musiman), dan seterusnya?

Untuk mengatasi permasalahan bisnis ini, teknik penambangan data telah diadopsi secara luas di seluruh sektor ritel online, ditambah dengan serangkaian metrik bisnis yang terkenal tentang profitabilitas dan nilai pelanggan, misalnya, model recency, frequency and fiscal (RFM), dan 3

model nilai kehidupan pelanggan. Bagi banyak pengecer daring di Inggris Raya dan internasional, terutama perusahaan terkemuka termasuk Amazon, Walmart, Tesco, Sainsbury's, Argos, Marks and Spencer, John Lewis, dan EasyJet, penambangan data kini telah menjadi praktik umum dan bagian integral dari proses bisnis dalam menciptakan intelijen bisnis yang berpusat pada pelanggan dan mendukung pemasaran yang berpusat pada pelanggan.

Meskipun banyak merek ritel daring ternama yang menerapkan teknik penambangan data sebagai alat penting untuk memperoleh keunggulan kompetitif di pasar, masih banyak merek-merek kecil dan pendatang baru yang ingin mempraktikkan pemasaran yang berpusat pada konsumen tetapi secara teknis tidak memiliki pengetahuan dan keahlian yang diperlukan untuk melakukannya.

Dalam artikel ini studi kasus penggunaan data teknik penambangan dalam intelijen bisnis yang berpusat pada pelanggan untuk pengecer daring disajikan. Pengecer daring yang dipertimbangkan di sini adalah pengecer yang umum: bisnis kecil dan pendatang baru yang relatif baru di sektor ritel daring, yang menyadari pentingnya menjadi analitis dalam bisnis daring dan teknik penambangan data saat ini, namun, kurang memiliki kesadaran teknis dan sumber daya.

Tujuan utama dari analisis ini adalah untuk membantu bisnis lebih memahami pelanggannya dan karena itu melakukan perilaku yang berpusat pada pelanggan pemasaran yang lebih efektif. Berdasarkan model RFM, pelanggan bisnis telah disegmentasikan ke dalam berbagai kelompok yang bermakna menggunakan algoritma pengelompokan k-means dan induksi pohon keputusan, dan karakteristik utama konsumen di setiap segmen telah diidentifikasi dengan jelas. Oleh karena itu, serangkaian rekomendasi diberikan kepada



bisnis pada pemasaran yang berpusat pada pelanggan dan tugas analisis data lebih lanjut. Analisis dikembangkan secara bertahap. SAS Enterprise Guide dan SAS Enterprise 6 – 9

Penambang telah dipekerjakan dalam penelitian ini.

Sisa artikel ini disusun sebagai berikut. Bagian selanjutnya memberikan informasi latar belakang tentang pengecer daring yang dipelajari dalam artikel beserta kumpulan data terkait yang akan dieksplorasi. Bagian selanjutnya membahas secara terperinci

tentang langkah-langkah dan tugas utama untuk pra-pemrosesan data guna membuat kumpulan data target yang sesuai untuk analisis lebih lanjut yang diperlukan. Pada bagian berikutnya, analisis pengelompokan k-means dilakukan dan serangkaian kluster dan segmen yang bermakna dari kumpulan data target telah diidentifikasi. Pembahasan terperinci tentang masing-masing

cluster diberikan, dan segmentasi disempurnakan lebih lanjut dengan menggunakan induksi pohon keputusan. Bagian kedua terakhir merangkum inti dari konsumen yang berpusat pada

intelijen bisnis berdasarkan hasil analisis, dan memberikan beberapa rekomendasi konkret kepada pengecer online

bertujuan untuk memaksimalkan keuntungan bagi bisnis. Akhirnya, pernyataan penutup diberikan di bagian terakhir.

LATAR BELAKANG BISNIS DAN DATA TERKAIT

Pengecer online yang dipertimbangkan dalam artikel ini berbasis di Inggris dan terdaftar

bisnis non-toko dengan sekitar 80 anggota

staf. Perusahaan ini didirikan pada tahun 1981 terutama menjual hadiah unik untuk segala acara.

Selama bertahun-tahun di masa lalu, pedagang sangat bergantung pada katalog surat langsung, dan pesanan diterima melalui panggilan telepon. Baru 2 tahun yang lalu perusahaan meluncurkan situs webnya sendiri dan beralih

sepenuhnya ke Web. Sejak saat itu perusahaan telah mempertahankan jumlah pelanggan yang stabil dan sehat dari seluruh bagian Inggris Raya dan Eropa, dan telah mengumpulkan sejumlah besar data tentang banyak pelanggan. Perusahaan juga menggunakan Amazon.co.uk untuk memasarkan dan menjual

produk.

Dataset transaksi pelanggan yang dimiliki oleh pedagang memiliki 11 variabel seperti yang ditunjukkan pada Tabel 1 dan berisi semua transaksi

terjadi pada tahun 2010 dan 2011. Perlu dicatat bahwa variabel *PostCode* adalah

penting bagi bisnis karena menyediakan informasi penting yang membuat setiap individu

konsumen dapat dikenali dan dilacak, sehingga memungkinkan dilakukannya beberapa analisis mendalam dalam studi ini.

Sebagai studi percontohan pertama bagi bisnis untuk menghasilkan kecerdasan pelanggan yang masuk akal, hanya transaksi yang dibuat dari 1 Januari 2011 hingga 31 Desember 2011 yang dieksplorasi dalam artikel ini. Selama periode tertentu, terdapat total 22.190 transaksi yang valid, yang terkait dengan 4.381 kode pos berbeda yang valid. Sesuai dengan

Tabel 1 : Variabel dalam dataset transaksi pelanggan (4381 contoh)

<i>Nama variabel</i>	<i>Tipe data</i>	<i>Deskripsi; nilai dan makna khas</i>
Faktur	Nominal	Nomor faktur; nomor integral 6 digit yang ditetapkan secara unik untuk setiap transaksi
Kode Saham	Nominal	Kode produk (item); nomor integral 5 digit yang ditetapkan secara unik untuk setiap produk yang berbeda Nama
Keterangan	Nominal	produk (item); KARTU I LOVE LONDON Jumlah setiap
Kuantitas	numerik	produk (item) per transaksi Harga produk per unit dalam
Harga	numerik	poundsterling; £45,23 Hari dan waktu setiap
Tanggal Faktur	numerik	transaksi dibuat; 31/05/2011 15:59 Alamat pengiriman baris 1; 103 Borough Road Alamat
Alamat Baris 1	Nominal	pengiriman baris 2; Elephant and Castle Alamat
Alamat Baris 2	Nominal	pengiriman baris 3; London Kode pos alamat pengiriman,
Alamat Baris 3	Nominal	terutama untuk konsumen dari Inggris
Kode Pos	Nominal	Raya; SE1 0AA Negara alamat pengiriman; Inggris
Negara	Nominal	



transaksi ini, terdapat 406.830 contoh (baris rekaman) dalam kumpulan data, masing-masing untuk item tertentu yang terdapat dalam transaksi.

Rata-rata, setiap kode pos dikaitkan dengan lima transaksi, yaitu, setiap pelanggan telah membeli produk dari pengecer daring sekitar sekali setiap 2 bulan. Selain itu, hanya konsumen dari Inggris Raya yang dianalisis.

Menarik untuk diperhatikan bahwa rata-rata jumlah produk (item) berbeda yang terdapat dalam setiap transaksi yang terjadi pada tahun 2011 adalah 18,3 (= 406.830 / 22.190). Hal ini tampaknya menunjukkan bahwa banyak konsumen bisnis tersebut pelanggan organisasi, bukan pelanggan individu.

PRA-PEMROSESAN DATA Untuk melakukan analisis pengelompokan berbasis model RFM yang diperlukan, kumpulan data asli perlu diproses terlebih dahulu. Langkah-langkah utama dan tugas-tugas relevan yang terlibat dalam persiapan data adalah sebagai berikut:

1. Pilih variabel yang sesuai dari kumpulan data yang diberikan. Dalam kasus kami, enam variabel berikut telah dipilih: *Faktur*, *Kode Stok*, *Kuantitas*, *Harga*, *Tanggal Faktur*, dan *Kode Pos*.
2. Buat variabel agregat bernama *Jumlah*, dengan mengalikan *Kuantitas* dengan *Harga*, yang memberikan jumlah total uang yang dibelanjakan per produk/item dalam setiap transaksi.
3. Pisahkan variabel *InvoiceDate* menjadi dua variabel *Tanggal* dan *Waktu*. Ini memungkinkan transaksi berbeda yang dibuat oleh konsumen yang sama pada hari yang sama tetapi pada waktu yang berbeda harus diperlakukan secara terpisah.
4. Saring semua transaksi yang tidak memiliki kode pos terkait. Ini menyelesaikan masalah nilai yang hilang terkait variabel *PostCode*.

Selain itu, saring semua transaksi yang tidak terkait dengan kode pos Inggris Raya.

5. Urutkan dataset berdasarkan *Kode Pos* dan membuat tiga variabel agregat penting yaitu *Keterkinian*, *Frekuensi* dan *Moneter*. Hitunglah nilai variabel berikut ini per kode pos.

Dengan mengikuti langkah-langkah ini, kumpulan data target untuk analisis telah dibuat. Kumpulan data asli dalam format MS Excel, dan diubah menjadi kumpulan data target akhir dalam format SAS di SAS Enterprise Guide 4.2.

Bagian dari dataset target ditunjukkan pada Gambar 1 dan variabel dalam dataset target dan statistiknya dijelaskan

pada Tabel 2 dan 3. Prosedur SAS *proc means* dan *proc sql* digunakan untuk mentransformasi dataset dan menghitung

nilai untuk variabel *Recency*, *Frequency* dan *Moneter*, untuk setiap kode pos yang diberikan, masing-masing. Sebagai contoh, Tabel 4 memberikan kode SAS relevan yang digunakan untuk menghitung nilai untuk *Moneter*. Akhirnya, kumpulan data target diunggah ke SAS Enterprise Miner 6.2 untuk dianalisis.

MODEL RFM BERBASIS ANALISIS PENGELOMPOKAN

Pengelompokan

Dengan kumpulan data target yang telah disiapkan, kami bermaksud mengidentifikasi apakah konsumen dapat disegmentasikan secara bermakna berdasarkan nilai kebaruan, frekuensi, dan moneter. Algoritma pengelompokan k-means digunakan untuk tujuan ini, dan dapat dilakukan dengan mudah menggunakan simpul Cluster di SAS Enterprise Miner.

Seperti yang diketahui, pengelompokan k-means Algoritma ini sangat sensitif terhadap kumpulan data yang berisi outlier (anomali) atau variabel yang skala atau besarnya tidak dapat dibandingkan. Pemeriksaan histogram variabel *Recency*, *Frequency* dan *Monetary*

dari dataset target di SAS Enterprise Miner, seperti yang diilustrasikan pada Gambar 2, terlihat dia jelas bahwa ada beberapa contoh memiliki perbedaan moneter dan

Filter and Sort Query Builder Data Describe Graph Analyze Export Send To										
	Buyer	First_Purchase	Recency	Frequency	Monetary	Min	Max	Mean		
2392	XX18 4ES	12	12	1	306.84	1.25	30	7.14		
2393	XX2 4HG	4	4	1	1487.6	237.6	1250	743.80		
2394	XX2 5XX	5	0	5	1952.45	0.39	76.32	22.19		
2395	XX20 6AB	11	1	2	483.26	0.84	17.34	4.56		
2396	XX20 6AQ	7	7	1	157.9	10.5	75	26.32		
2397	XX20 6HJ	3	1	3	541.46	1.65	59.4	15.93		
2398	XX21 8YN	1	1	1	167.62	0.19	15.12	3.64		
2399	XX22 3RB	3	1	3	454.01	2.9	25.5	9.27		
2400	XX24 8SA	11	11	1	344.14	1.45	40.56	11.87		
2401	XX25 5HU	10	2	4	1498.51	0.12	45	5.01		
2402	XX26 1AG	12	3	4	226.75	13.05	26.1	18.90		
2403	XX26 3QN	8	8	1	185.65	15	35.7	23.21		
2404	XX30 6AP	12	8	3	215.72	0.58	47.7	4.79		
2405	XX30 6BJ	8	8	1	175.27	0.65	60	6.04		
2406	XX31 7AD	11	4	4	1133.07	0.29	41.5	5.29		
2407	XX31 7JF	8	4	3	1240.2	9.12	264	53.92		
2408	XX31 7JN	8	2	4	2404.17	0.39	74.25	9.50		
2409	XX33 0EA	9	3	2	993.18	8.5	39.6	19.86		
2410	XX33 0EN	1	1	2	388.79	0.39	34.68	4.68		
2411	XX33 0QL	12	1	7	2827.93	0.29	142.8	10.47		
2412	XX34 1HA	7	7	1	2044.37	5.04	207.5	37.17		
2413	XX34 2DS	1	1	1	161.67	2.5	19.8	12.44		
2414	XX34 2HA	8	8	1	116.01	2.1	30	12.89		
2415	XX34 3OA	11	7	2	309.36	1.1	17.4	4.91		
2416	XX37 6QU	1	1	2	193.42	0.19	11.8	3.28		
2417	XX39 4HL	2	2	1	390.07	0.42	18.72	3.90		

Gambar 1 : Contoh dataset target.

Tabel 2 : Variabel dalam dataset target

Nama variabel Tipe data Deskripsi

Pembeli	Nominal Sesuai dengan setiap kode pos yang berbeda
Kebaruan	Keterkinian Numerik dalam bulan
Pembelian_Pertama	Numerik Waktu dalam bulan sejak pembelian pertama pada tahun 2011
Frekuensi	Frekuensi Numerik Pembelian per Kode Pos
Moneter	Numerik Jumlah total yang dibelanjakan per kode pos
Minimum	Pengeluaran minimum numerik per kode pos
Maksimum	Angka Pengeluaran maksimum per kode pos
Berarti	Median pengeluaran numerik per kode pos

nilai frekuensi dibandingkan dengan mayoritas contoh dalam kumpulan data. Ini

contoh-contoh tersebut valid dari sudut pandang bisnis karena merupakan catatan transaksi yang asli;

namun, contoh-contoh tersebut merupakan outlier dari sudut pandang analisis data. Oleh karena itu, contoh-contoh tersebut harus dipisahkan dari

mayoritas dan diperlakukan secara terpisah. Selain itu, ketiga variabel tersebut tidak berada pada skala yang sebanding, dan rentang nilainya cukup

Tabel 3: Ringkasan kumpulan data target (3799 contoh)

Nama variabel Minimum Median Maksimum

Kebaruan	0	3,2 12 4,9 169
Frekuensi		1586,63 88 125,38
Moneter	1 3.75	7,5 12
Pembelian_Pertama	0	

Tabel 4 : Contoh kode SAS untuk menghitung nilai mata uang

proc berarti data= NamaPerpustakaanAnda.

SortedOriginalDataset n jumlah min maks rata-rata; var

Jumlah; menurut

Kode Pos; keluaran

keluar= NamaPerpustakaanAnda.TagretDatasetMonetary

(drop=_type_ _freq_) n=n jumlah=jumlah min=min maks=maks

mean=rata-rata;

lari;

berbeda: Keterkinian [0,12]; Frekuensi [1,169] dan Moneter [3,88 125]. Oleh karena itu, variabel-variabel ini harus dinormalisasi sebelum analisis pengelompokan.

Berdasarkan pemahaman awal tentang dataset, diagram proyek telah disiapkan



Gambar 2: Distribusi variabel Keterkinian, Frekuensi dan Moneter.



Gambar 3 : Diagram proyek di SAS Enterprise Miner 6.2.

Tabel 5: Ringkasan kumpulan data target yang difilter (3726 contoh)

Variabel nama	Minimum	Median	Maksimum
Kebaruan	0	3.2	12
Frekuensi		4.1	28
Moneter	1 3.75	1565,70	13 110,02 7,5
Pembelian_Pertama	0		12

di SAS Enterprise Miner untuk analisis pengelompokan seperti yang digambarkan pada Gambar 3. Ada empat node dalam diagram. Di node Sumber Data (Target Dataset), tiga variabel *Keterkinian*, *Frekuensi*, dan *Moneter* dipilih sebagai input untuk analisis pengelompokan. Node Filter ditetapkan untuk mengecualikan dari analisis setiap contoh yang memiliki nilai langka untuk setiap variabel yang terlibat, dan nilai batas minimum untuk nilai langka ditetapkan menjadi 1 persen dari jumlah total contoh yang dipertimbangkan. Misalnya, dari total 3799 contoh, hanya ada satu contoh yang mengambil nilai moneter lebih dari £ 87.684, dan oleh karena itu, contoh tersebut dikeluarkan dari analisis.

Secara keseluruhan terdapat 73 contoh yang dikecualikan oleh node Filter, dan ringkasan kumpulan data target yang difilter yang dihasilkan diberikan dalam Tabel 5. Pada node Cluster, transformasi rentang standar untuk normalisasi digunakan dengan jumlah cluster yang ditetapkan masing-masing 3, 4, dan 5, dan akhirnya, node Profil Segmen digunakan untuk membantu menafsirkan setiap cluster yang ditemukan.

Hasil pengelompokan dan segmen dengan Lima kluster ditunjukkan pada Tabel 6 dan 7 dan distribusi contoh dalam setiap kluster dirinci pada Gambar 4 dan 5. Segmentasi dengan lima kluster ini tampaknya memiliki interpretasi yang lebih jelas terhadap kumpulan data target daripada segmentasi dengan tiga dan empat kluster.

Tabel 6 : Contoh pada setiap cluster

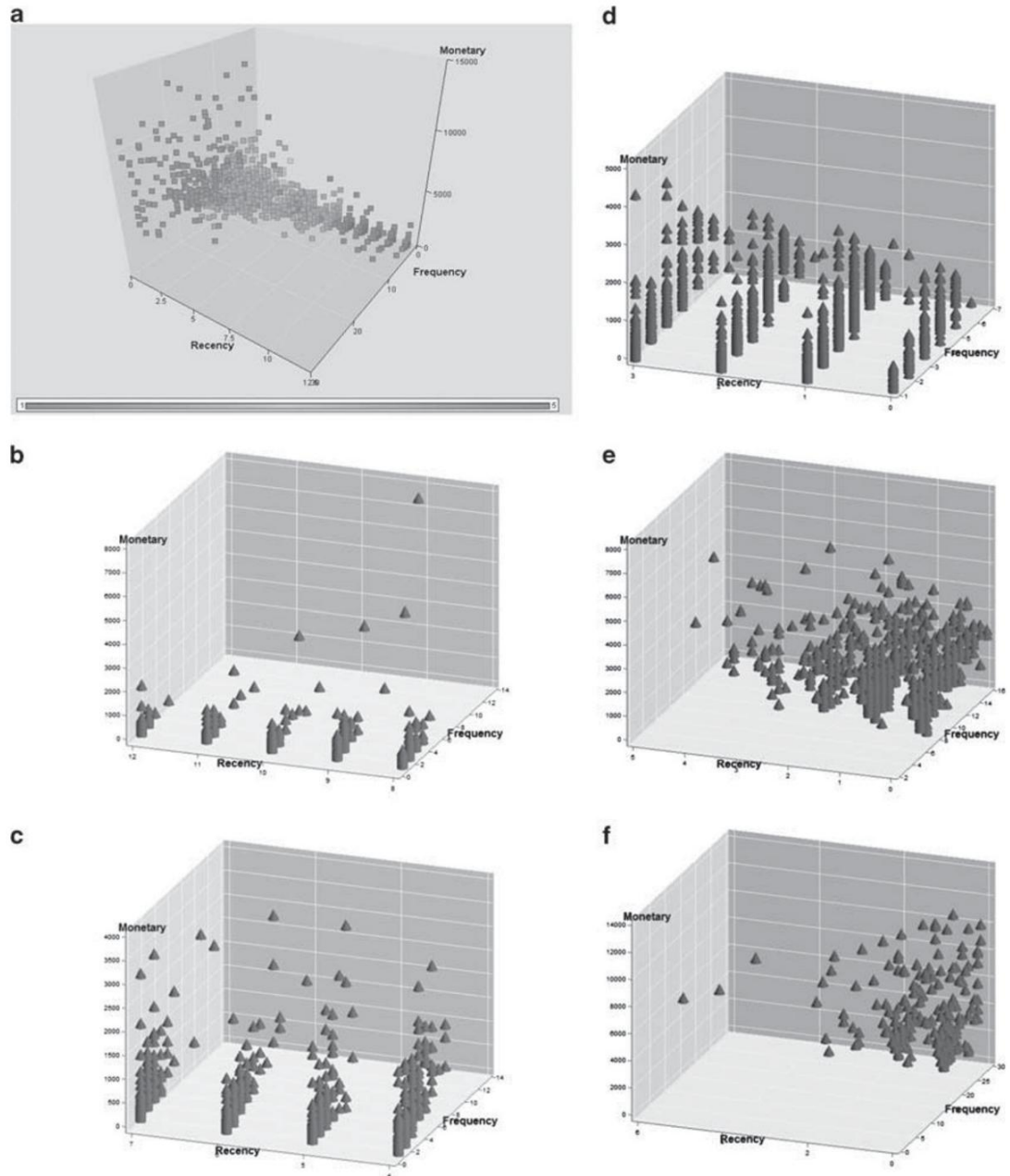
Gugus	Frekuensi kluster	Persentase
	527	14.14
	636	17.07
	1748	46.91
	627	16.83
	188	5.05

Tabel 7 : Statistik masing-masing kluster

	Minimum	Median	Maksimum
Kluster 1			
Kebaruan	8	9,8	12
Frekuensi		1,3	4
Moneter	1 3.75	361,20 7741,47	11,1 12
Pembelian_Pertama	8		
Gugus 2			
Kebaruan	4	5,4 7 2,3 13	
Frekuensi		586,19 3906,27	
Moneter	1	7,7 12	
Pembelian_Pertama	15 4		
Gugus 3			
Kebaruan	0	1,5	
Frekuensi		2,6	
Moneter	1 20.8	685,71	3 7
Pembelian_Pertama	0	5,3	4314.72 12
Gugus 4			
Kebaruan	0	1.0	
Frekuensi	3	8.3	5 16
Moneter	191.17 2425.09	7330.8 1.0	
Pembelian_Pertama	1		12
Gugus 5			
Kebaruan	0	0,7	
Frekuensi	3	17,7	6 28
Moneter	1641,48 5962,85	13 110,02 11,1	
Pembelian_Pertama	---		12

Memahami kluster Menafsirkan dan memahami setiap kluster yang diidentifikasi sangat penting dalam menghasilkan intelijen bisnis yang berpusat pada pelanggan.

Memeriksa Tabel 7 dan Gambar 4 dan 5 Menarik untuk melihat bahwa setiap kluster memang berisi sekelompok konsumen yang memiliki fitur unik dan hakiki tertentu sebagaimana dirinci di bawah ini.

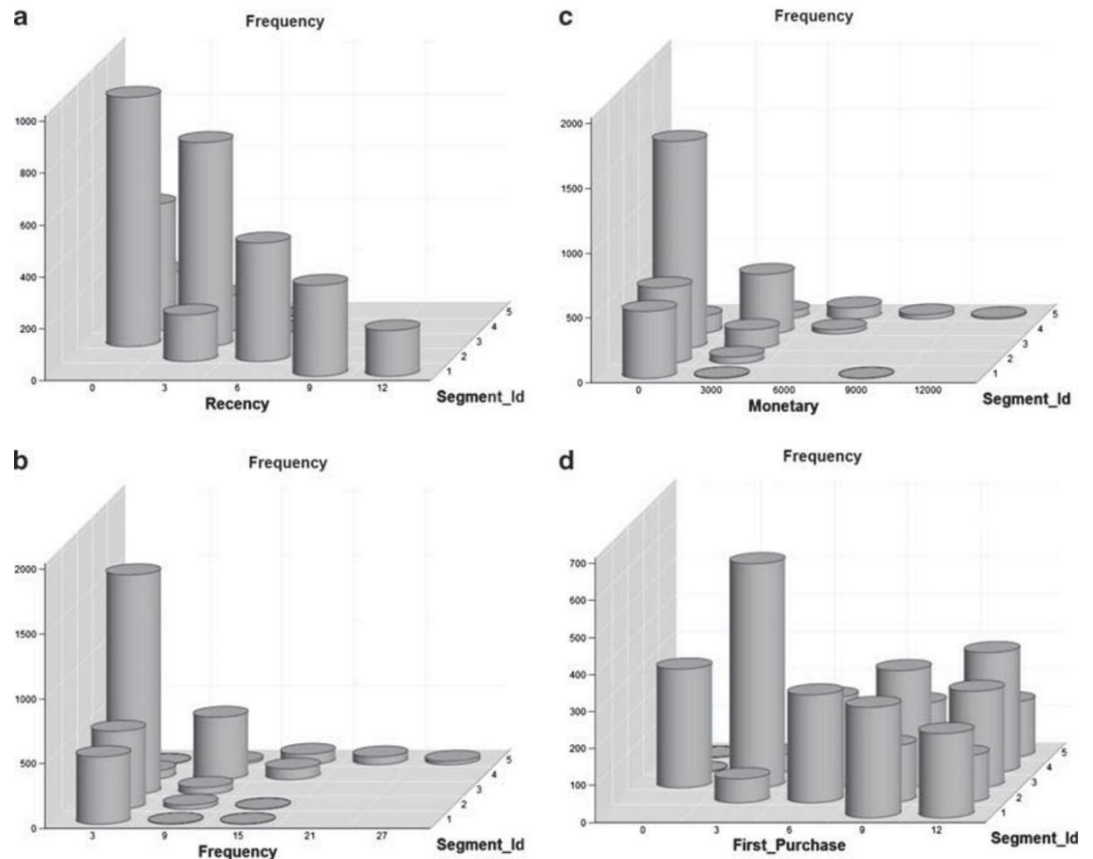


Gambar 4 : (a) Distribusi semua contoh yang diwarnai untuk klaster yang berbeda. (b) Distribusi contoh dalam klaster 1. (c) Distribusi contoh dalam klaster 2. (d) Distribusi contoh dalam klaster 3. (e) Distribusi contoh dalam klaster 4. (f) Distribusi contoh dalam klaster 5.

Klaster 1 mencakup sekitar 527 konsumen, terdiri dari 14,4 persen dari seluruh populasi. Kelompok ini tampaknya merupakan kelompok yang paling tidak menguntungkan karena tidak ada pelanggan dalam kelompok ini yang membeli apa pun pada paruh kedua tahun ini. Bahkan untuk

paruh pertama tahun ini, konsumen tidak sering berbelanja dan nilai rata-rata frekuensi hanya 1,3.

Dibandingkan dengan pelanggan di klaster 1, 188 pelanggan di klaster 5 sebagian besar memulai belanja melalui online



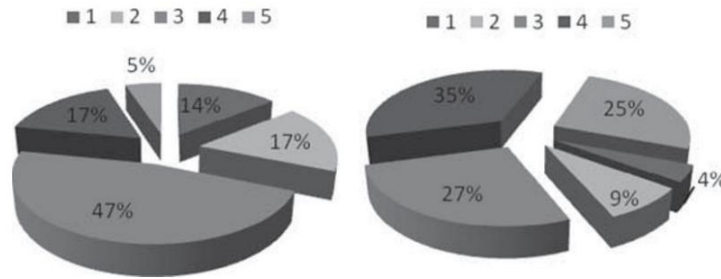
Gambar 5 : (a) Distribusi keterkinian berdasarkan klaster. (b) Distribusi frekuensi berdasarkan klaster. (c) Distribusi moneter berdasarkan klaster. (d) Distribusi pembelian pertama berdasarkan klaster.

pengecer di awal tahun, dan berlanjut hingga akhir tahun dengan nilai rata-rata recency 0,7. Mereka cukup sering membeli dan akibatnya, menghabiskan jumlah uang yang cukup tinggi. Kelompok konsumen ini dapat dikategorikan sebagai very high recency, very high frequency dan very high money dengan pengeluaran per konsumen yang tinggi. Bahkan, 188 konsumen tersebut menyumbang 25,5 persen dari total penjualan pada tahun tersebut. Kelompok ini, meskipun yang terkecil (hanya terdiri dari 5,05 persen dari seluruh populasi), tampaknya merupakan kelompok yang paling menguntungkan.

Klaster 4 berisi sekitar 627 konsumen dengan nilai frekuensi dan moneter yang sangat tinggi, meskipun lebih rendah dibandingkan klaster 5. Kelompok ini tampaknya merupakan kelompok dengan keuntungan tertinggi kedua.

Ada sekitar 459 konsumen dalam klaster 2. Dibandingkan dengan kelompok 4 dan 5, kelompok pelanggan ini memiliki frekuensi yang lebih rendah sepanjang tahun dan nilai rata-rata moneter yang jauh lebih kecil, yang menunjukkan jumlah pengeluaran per konsumen yang jauh lebih kecil. Kelompok ini dapat dikategorikan sebagai low recency, high frequency, dan medium moneter dengan pengeluaran sedang per konsumen.

Klaster 3 merupakan kelompok dengan jumlah konsumen terbesar, yaitu 1748 konsumen. Konsumen dalam kelompok ini memiliki nilai frekuensi yang wajar. Dibandingkan dengan klaster 2 dan 4, kelompok ini memiliki nilai moneter yang lebih rendah namun wajar karena kelompok ini mencakup banyak konsumen yang baru terdaftar dan mulai berbelanja dengan pengecer baru-baru ini. Kelompok ini tampaknya mewakili konsumen biasa.



Gambar 6: Segmentasi pelanggan (kiri) dan penjualan terkait (kanan) berdasarkan kluster.

konsumen dan karena itu memiliki tingkat tertentu ketidakpastian dalam hal keuntungan. Dalam pandangan jangka panjang, beberapa konsumen mungkin berpotensi sangat menguntungkan atau tidak menguntungkan sama sekali.

Kami menggunakan Gambar 6 untuk meringkas Analisis yang telah dilakukan sejauh ini: dari keseluruhan populasi konsumen, 47 persen di antaranya adalah pembeli biasa dengan pengeluaran dan frekuensi yang wajar, sekitar 34 persen adalah pembeli dengan keuntungan sedang hingga tinggi, 5 persen adalah pembeli dengan keuntungan sangat tinggi, dan 14 persen sisanya adalah pembeli dengan keuntungan sangat rendah. Sekitar 22 persen konsumen menyumbang sekitar 60 persen dari total penjualan. Secara keseluruhan, bisnis tersebut tampaknya cukup sehat dalam hal keuntungan.

Meningkatkan analisis pengelompokan menggunakan pohon keputusan

Seperti yang telah dibahas di atas, kluster 3 merupakan kluster yang paling beragam di antara lima kluster yang telah diidentifikasi, karena kluster ini mengandung kedua jenis pelanggan baru terdaftar dan pelanggan lama.

Untuk menyempurnakan segmentasi instansi dalam kluster ini, pohon keputusan telah digunakan untuk membuat beberapa segmen bersarang secara internal di dalam kluster, seperti ditunjukkan pada Gambar 5.

Dengan kata lain, segmen-segmen bersarang ini membentuk beberapa sub-kluster di dalam kluster 3, dan memungkinkan untuk mengkategorikan konsumen yang bersangkutan ke dalam beberapa kategori yang masuk akal. sub-kategori. Misalnya, seperti yang ditunjukkan pada Gambar 7 pelanggan dapat dibagi menjadi beberapa kategori seperti frekuensi lebih dari

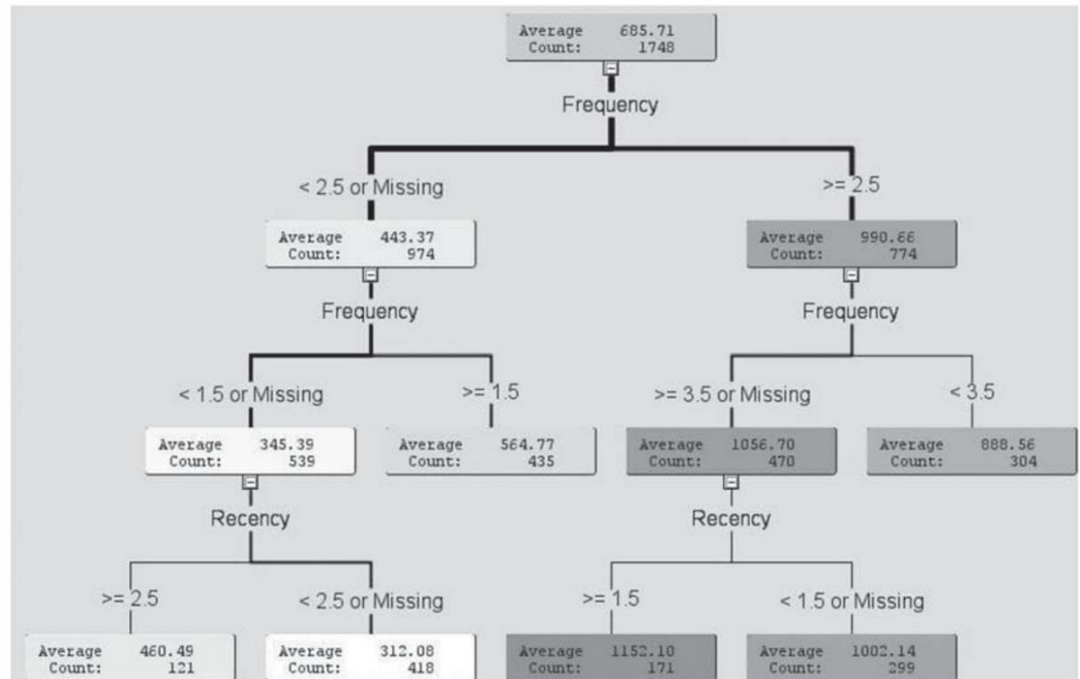
2,5 dengan nilai moneter rata-rata 990,66; dan frekuensi lebih dari 2,5 dan kurang dari 3,5 dengan nilai moneter rata-rata 1056,70 dan seterusnya. Selain itu, menarik untuk dicatat bahwa hubungan antara frekuensi dan moneter tampaknya merupakan hubungan linier monotonik.

BERPUSAT PADA PELANGGAN KECERDASAN BISNIS DAN REKOMENDASI

Konsumen paling berharga di dunia bisnis telah memberikan kontribusi lebih dari 60 persen dari total penjualan pada tahun 2011, sedangkan yang paling tidak bernilai hanya mencapai 4 persen dari total penjualan. Untuk masing-masing kelompok konsumen ini, penting untuk mengetahui lebih lanjut produk mana yang telah dibeli oleh pelanggan di setiap kelompok, produk mana yang paling sering dibeli bersama-sama, dan dalam urutan apa produk tersebut dibeli.

Bisnis dapat memperoleh pemahaman yang lebih baik tentang konsumen dengan mengeksplorasi hubungan antara kelompok konsumen dan produk yang telah mereka beli. Hubungan tersebut dapat diteliti pada tingkat produk/item dan juga pada tingkat kategori produk.

Banyak konsumen bisnis adalah konsumen organisasi dengan kuantitas produk yang tinggi per transaksi. Memeriksa pada waktu-waktu tertentu (musim), produk apa dan jenis produk apa yang sering mereka beli akan memberikan manfaat bagi bisnis. Ini juga akan



Gambar 7 : Segmentasi yang disempurnakan dari instans di kluster 3 menggunakan induksi pohon keputusan.

Menarik untuk melihat apakah ada perbedaan antara berbagai jenis pelanggan, yaitu pelanggan organisasi dan individu, dalam hal pola belanja mereka.

Memantau keberagaman kelompok pelanggan yang paling beragam dan memperkirakan pelanggan mana yang berpotensi berafiliasi dengan kelompok yang paling atau paling tidak menguntungkan sangat berguna bagi bisnis dalam jangka panjang. Mengidentifikasi prediktor atau indikator yang tepat untuk prediksi tersebut sangat berharga.

Aspek lain yang layak diselidiki lebih lanjut adalah menghubungkan kelompok konsumen dengan lokasi geografis. Korelasi ini, jika ada, dapat membantu bisnis untuk mempertimbangkan faktor-faktor lain, seperti budaya, adat istiadat, dan ekonomi, yang dapat memengaruhi minat dan preferensi pembelian konsumen.

KESIMPULAN

Sebuah studi kasus telah disajikan dalam artikel ini untuk menunjukkan bagaimana pendekatan yang berpusat pada pelanggan intelijen bisnis untuk pengecer online dapat

dapat dibuat dengan menggunakan teknik penambangan data. Kelompok pelanggan yang berbeda yang dicirikan dalam studi kasus dapat membantu bisnis lebih memahami pelanggannya dalam hal profitabilitasnya, dan karenanya, mengadopsi strategi pemasaran yang tepat untuk konsumen yang berbeda-beda.

Telah ditunjukkan dalam analisis ini bahwa ada dua langkah dalam keseluruhan proses penambangan data yang sangat penting dan paling memakan waktu: persiapan data dan interpretasi serta evaluasi model.

Penelitian lebih lanjut untuk bisnis ini meliputi: melakukan analisis asosiasi untuk menetapkan pola pembelian pelanggan sehubungan dengan produk mana yang sering dibeli bersama oleh pelanggan mana dan kelompok pelanggan mana; meningkatkan situs web pedagang untuk memungkinkan konsumen

aktivitas belanja dapat ditangkap dan dilacak secara instan dan akurat; dan memprediksi setiap nilai pelanggan untuk mengukus siklus hidup tingkat keberagaman setiap pelanggan.



UCAPAN TERIMA KASIH Penulis

mengucapkan terima kasih kepada para peninjau anonim atas komentar dan saran berharga mereka untuk meningkatkan kualitas artikel ini.

REFERENSI 1

Interactive Media in Retail Group (IMRG) . (2012)

Arsip pers, <http://www.imrg.com>, diakses Januari

Tahun 2012 .

2. Kumar V. dan Reinartz W. J. . (2006) *Manajemen Hubungan Pelanggan: Pendekatan Basis Data* , Hoboken, NJ: John Wiley & Sons.
- 3 Hughes, yang A. M. (2012) *Pemasaran Basis Data Strategis 4e: Rencana Induk untuk Memulai dan Mengelola Tabel Keuntungan, Program Pemasaran Berbasis Pelanggan*, McGraw-Hill Professional, AS.
- 4 Davenport, T dari H. (2009) *Mewujudkan Potensi Retail Analytics: Banyak Makanan untuk Mereka yang Memiliki* Appetite. Laporan Pengetahuan Kerja, Babson Executive Education.
5. Fuloria S. (2011) *Bagaimana Analisis Tingkat Lanjut Akan Menginformasikan dan Mengubah Ritel AS*. Laporan Cognizant, Juli, <http://www.cognizant.com/InsightsWhitepapers/How-Advanced-Analytics-Will-Inform-and-Transform-US-Retail.pdf> , diakses Januari 2012.
- 6 Collica R . S . (2007) *Segmentasi CRM dan Pengelompokan Menggunakan SAS Enterprise Miner*, Cary, NC: SAS Institute.
7. Cerrito, Meksiko P. B. (2007) *Pengantar Penambangan Data Menggunakan SAS Enterprise Miner*. Cary, NC: SAS Institute.
- 8 Sarma SAS —, S. (2007) *Pemodelan Prediktif dengan Enterprise Miner*. Cary, NC: Institut SAS.
- 9 Thompson , W. (2008) *Memahami Pelanggan Anda: Teknik Segmentasi untuk Mendapatkan Wawasan Pelanggan dan Memprediksi Risiko dalam Industri Telekomunikasi*. Makalah 154-2008, SAS Global Forum, 16 – 19 Maret, San Antonio, TX.