

PENERAPAN ALGORITMA C4.5 PADA ANALISIS DATA PENJUALAN OBAT HERBAL UNTUK PREDIKSI PENJUALAN DI TOKO ONLINE XXX

Oleh: Chaerul Anwar¹, Feny Erie Emawati²

Program Studi Sistem Informasi^{1,2}

Universitas Pembangunan Jaya^{1,2}

Email: chaerul.anwar@upj.ac.id¹, fenevscnc@gmail.com²

Abstrak

Obat herbal atau ramuan tradisional merupakan suatu metode pengobatan alternatif dengan menggunakan tanaman dan bahan baku alamiah yang tidak mengandung bahan kimia yang dapat menimbulkan berbagai efek samping seperti obat yang diberikan oleh dokter secara medis. Toko online XXX merupakan salah satu distributor obat herbal yang menjual beberapa jenis obat minum dengan khasiat yang berbeda-beda. Adapun penelitian ini bertujuan untuk menganalisis penjualan yang telah terjadi serta memprediksi penjualan obat herbal pada masa mendatang menggunakan algoritma C4.5 yang berbentuk pohon keputusan. Hasil analisis menunjukkan bahwa pohon keputusan yang dihasilkan dari perhitungan nilai entropy dan gain pada beberapa sampel atribut yaitu Paket Perawatan Wajah dan Sabun Pembersih Wajah memiliki variasi yang berbeda. Fakta membuktikan bahwa semakin tinggi nilai entropy yang didapatkan dalam perhitungan, maka semakin sedikit pohon keputusan yang dihasilkan. Begitu pula sebaliknya.

Kata kunci: Algoritma C4.5, Analisis Data Penjualan, Prediksi.

PENDAHULUAN

Toko online XXX merupakan distributor obat herbal yang melakukan penjualan secara daring dengan memasarkan produk-produk melalui laman iklan di Google, media sosial seperti Facebook dan Instagram serta melalui laman perdagangan elektronik atau e-Commerce seperti Tokopedia, Shopee, Blibli dan Lazada. Pada toko online XXX, belum pernah dilakukan analisis pada data penjualan secara manual maupun menggunakan sistem, yang bertujuan untuk memprediksi penjualan pada masa yang akan datang. Prediksi yang dihasilkan dapat meminimalisir dan bahkan mencegah penyediaan stok produk yang berlebihan karena produk dapat rusak karena kadaluarsa. Selain itu, dengan prediksi penjualan ini, kurangnya stok yang tersedia juga dapat dihindari agar tidak ada pelanggan yang kecewa pada saat ingin membeli namun ketersediaan stok malah kosong karena belum dilakukan prediksi penjualan untuk mengetahui perkiraan terdekat atas stok yang perlu disediakan. Kesalahan dalam menentukan stok dapat berakibat buruk dalam proses penjualan yang berlangsung di toko online XXX dan dapat meningkatkan kemungkinan kerugian yang akan ditanggung. Oleh karena itu, suatu analisis terhadap data penjualan untuk memprediksi penjualan pada masa mendatang dapat menjadi pilihan yang tepat.

LATAR BELAKANG

Secara umum, toko online ini menjalankan operasional yaitu memasarkan dan menjual berbagai produk pengobatan herbal yang meliputi kesehatan pencernaan, saraf, wajah dan juga rambut. Pada dasarnya, toko ini melakukan pengenalan produk melalui website iklan yang muncul pada mesin pencarian maupun jejaring sosial menyesuaikan dengan target dan kriteria pelanggan yang dituju. Selanjutnya apabila calon pelanggan

tertarik pada iklan yang tampil pada halaman yang sedang mereka buka, maka biasanya pelanggan akan langsung membuka halaman website iklan, lalu membaca berbagai rincian mengenai produk hingga pada bagian paling bawah yaitu tombol untuk menghubungi CS untuk menanyakan perihal produk, bagaimana cara melakukan pemesanan, pembayaran dan proses pengiriman pesanan yang akan dilakukan.

Penelitian ini berfokus pada analisis data penjualan obat herbal yang dipasarkan dan dijual oleh Toko Online XXX. Karena sebelumnya Toko Online XXX belum pernah melakukan analisis yang cukup terperinci untuk memenuhi kebutuhan penyediaan produk agar selalu seimbang dan tidak berlebihan maupun kekurangan. Namun analisis ini dilakukan bukan hanya untuk menyeimbangkan stok produk yang dimiliki, tetapi juga untuk memprediksi besarnya penjualan yang akan terjadi di masa mendatang. Sebelumnya ketersediaan atau stok produk tidak menyesuaikan pada penjualan yang ada sehingga produk beresiko menumpuk dan dapat menyebabkan produk dapat rusak karena waktu kadaluwarsa yang semakin dekat namun produk belum terjual atau terpakai. Hal tersebut terjadi karena target penjualan produk yang belum terpenuhi namun dipengaruhi juga oleh cakupan pemasaran yang dilakukan. Oleh karena itu, agar penjualan serta pemasarannya dapat berjalan secara efektif dan efisien, ada baiknya dilakukan analisis pada data penjualan yang sudah ada untuk memprediksi kesuksesan penjualan masa bulan berikutnya.

Salah satu solusi yang tepat dalam menjalankan analisis yang dimaksud adalah dengan menerapkan algoritma C4.5 di mana algoritma ini merupakan suatu konsep pemikiran berupa gambaran dengan bentuk pohon keputusan (decision tree) yang memiliki node atau titik internal untuk mendeskripsikan serta menghubungkan setiap atribut yang ada. Algoritma C4.5 juga merupakan metode yang dapat dikonversikan menjadi aturan klasifikasi (classification rule). Aturan klasifikasi atau classification rule merupakan salah satu metode turunan dari konsep machine learning. Pada penerapannya, antara satu atribut dengan atribut lainnya dikelompokkan ke dalam beberapa kategori. Setiap atribut yang berada dalam sebuah kategori memiliki sifat yang hampir sama. Oleh karena itu pengelompokkan dilakukan berdasarkan sifat suatu atribut yang familiar dengan atribut lainnya. Algoritma ini dikembangkan oleh Ross Quinlan yang merupakan perpanjangan atau pelebaran konsep dari algoritma ID3 sebelumnya. Algoritma ID3 sendiri merupakan konsep logika untuk menghasilkan pohon keputusan namun hanya terbatas pada data diskrit atau nominal.

RUMUSAN MASALAH

Rumusan masalah dari latar belakang adalah sebagai berikut : Bagaimana penggunaan Algoritma C4.5 dalam memprediksi stok penjualan pada toko online XXX.

TUJUAN PENELITIAN

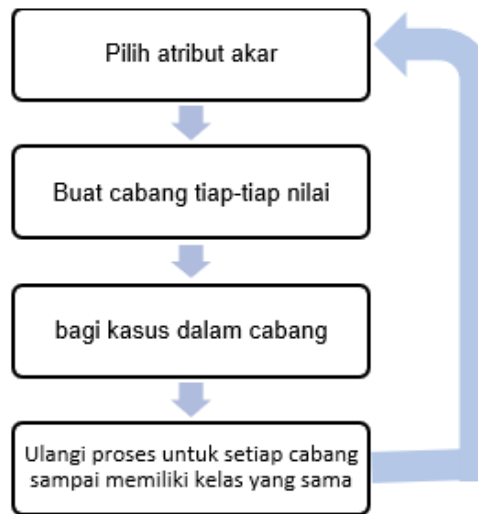
Tujuan penelitian merupakan jawaban atau tujuan yang ingin dicapai penulis dalam penelitian. Oleh karena itu, tujuan dari penelitian ini adalah:

1. Untuk mengetahui kebutuhan pengguna aplikasi yang seperti apa yang dibutuhkan pencatatan keuangan dan pencatatan aset karang taruna DIFTA 12.
2. Untuk mengetahui *task* apa saja yang diperlukan untuk memenuhi kebutuhan pencatatan keuangan dan aset agar terlaksana sebagaimana mestinya.
3. Untuk mengetahui proses bisnis apa saja dalam hal aktifitas pencatatan keuangan dan aset di karang taruna DIFTA 12.

METODE PENELITIAN

Metode yang digunakan pada analisis penjualan yang hanya memiliki satu atribut pada setiap record atau transaksinya seperti setiap transaksi yang dimiliki oleh Toko Online XXX cocok menerapkan algoritma C4.5 karena dapat digunakan untuk data diskrit atau nominal dalam jumlah yang tak terbatas dan jika pada masa mendatang

akan dilakukan analisis kembali pada atribut lain yang bukan merupakan data diskrit, masih dapat menerapkan algoritma ini. Meskipun terjadinya hal tersebut memiliki kemungkinan yang cukup kecil. Pada dasarnya, secara teknis terdapat langkah-langkah dalam menerapkan algoritma C4.5, di antaranya adalah sebagai berikut :



Gambar 1. Algoritma C4.5

Atribut dalam hal ini adalah masing-masing produk yang terdapat dalam transaksi, dengan nilai yang dimaksud adalah angka penjualan per produknya. Pada algoritma C4.5, terdapat nilai entropy dan gain yang akan mempengaruhi nilai akhir yang akan ditampilkan pada pohon keputusan. Sementara itu untuk menentukan nilai entropy dapat menggunakan rumus seperti yang tertera pada persamaan 1, di mana: S = Himpunan Kasus, n = Jumlah Partisi S, dan pi = Proporsi Si terhadap S. Lalu untuk mendapatkan nilai gain adalah dengan menggunakan rumus yang tertera pada persamaan 2, di mana: S = Himpunan Kasus, A = Atribut, n = Jumlah Partisi Atribut A, | Si | = Jumlah kasus pada partisi ke-l, dan | S | = Jumlah Kasus dalam S.

ANALISA DATA

Pada bagian pembahasan ini akan dijelaskan bagaimana proses perhitungan untuk mengetahui nilai entropy dan gain berdasarkan angka penjualan dan atribut atau produk seperti yang tertera pada Tabel 1, di mana fluktuasi penjualan setiap bulannya diindikasikan dengan N (Naik) dan T (Turun) jika dibandingkan dengan angka penjualan pada bulan sebelumnya. Analisis ini dilakukan pada data penjualan selama tiga bulan yaitu dari bulan Januari hingga Maret 2019 dengan 12 produk atau yang disebut atribut pada algoritma C4.5 dan akan digunakan sebagai salah satu penentu hasil angka entropy dan gain. Selain itu, banyaknya nilai N dan T juga mempengaruhi nilai akhir yang akan ditampilkan pada pohon keputusan.

Maka untuk mengetahui nilai entropy dapat dilakukan perhitungan yang dapat dilakukan sebagai berikut:

$$\begin{aligned}
 \text{entropy (S)} &= (- (4/12) * \text{Log}_2(4/12) + (- (4/12) * \text{Log}_2(4/12)) \\
 &= \mathbf{1,014930619}
 \end{aligned}$$

Tabel 1. Data Penjualan Januari s/d Maret 2019.

PRODUK	UKURAN	TERJUAL							TOTAL
		DES	JAN	KATEGORI	FEB	KATEGORI	MAR	KATEGORI	
PAKET									
PERAWATAN	-	0	233	N	185	T	90	T	508
WAJAH									
KRIM MALAM	100 GR	0	1	N	0	T	0	T	1
WAJAH									
KRIM SIANG	100 GR	0	0	-	2	N	0	T	2
WAJAH									
TONER WAJAH	100 GR	0	0	-	4	N	0	T	4
MADU ANAK	125 GR	0	3	N	0	T	57	N	60
MADU									
REMATIK	175 GR	274	223	T	107	T	85	T	415
MADU									
REMATIK	350 GR	363	378	N	208	T	144	T	730
MADU									
LAMBUNG	125 GR	268	498	N	559	N	893	N	1950
MADU									
PENCERNAAN	125 GR	40	26	T	25	T	22	T	73
MADU PAHIT	125GR	3	0	T	0	-	0	-	3
MINYAK									
KEMIRI	100 GR	0	3	N	1	T	0	T	4
SABUN									
PEMBERSIH	50GR	0	0	-	4	N	3	T	7
WAJAH									

Dari perhitungan tersebut dapat diperoleh nilai hasil dataset seperti yang tertera pada Tabel 2 yang merupakan **nilai entropy yang dihitung secara keseluruhan berdasarkan banyaknya nilai N dan T pada data penjualan.**

Tabel 2. Hasil Perhitungan Dataset.

Total Kasus	Sum (N)	Sum (T)	Entropy Total
36	12	19	1,014930619

Setelah diketahui nilai entropy dari keseluruhan jumlah partisi yaitu 12 atribut produk, perlu ditentukan juga nilai entropy dari setiap atribut produk menggunakan rumus yang sama. Salah satu contohnya adalah perhitungan nilai entropy dari kategori N pada angka penjualan bulan Februari

$$entropy(S) = (- (12/36) * \text{Log}_2(12/36) + (- (19/36) * \text{Log}_2(19/36)) = 0,528320834$$

Selanjutnya adalah proses perhitungan untuk mengetahui nilai gain berdasarkan dari keseluruhan atribut salah satunya menggunakan nilai entropy yang telah diketahui pada langkah sebelumnya sehingga perhitungan nilai gain nya adalah sebagai berikut

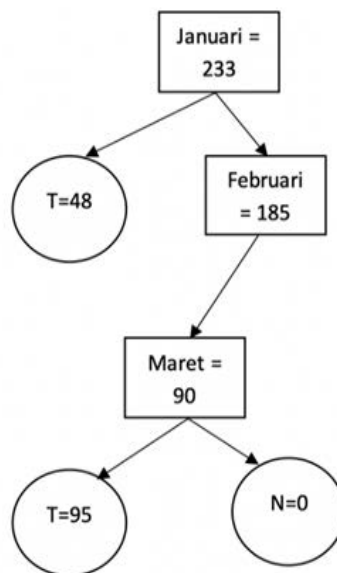
$$Gain(S, A) = 1,014930619 - ((4 \div 12 \times 0,528320834) + (7 \div 12 \times 0,453604421)) = 0,574221095$$

Berdasarkan hasil perhitungan nilai entropy dan gain dari masing-masing atribut, maka didapatkan nilai keduanya pada setiap bulan penjualan seperti yang tertera pada Tabel 3.

Tabel 3. Analisis Atribut, Nilai, Banyaknya Kejadian Nilai, Entropy dan Gain

No	Nilai	Jan	Entropy	Feb	Entropy	Mar	Entropy
1	Sum (N)	6	0,5	4	0,528320834	2	0,430827083
2	Sum (T)	3	0,5	7	0,453604421	9	0,311278124
	Entropy	1		0,981925254		0,742105208	
	Gain	0,639930619		0,574221095		0,709667512	

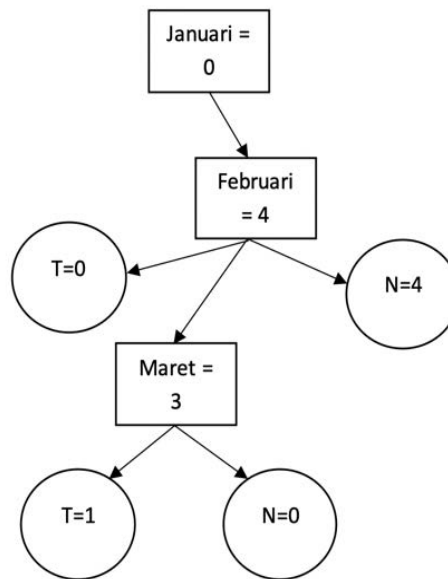
Apabila nilai entropy dan gain sudah didapatkan berdasarkan pengkategorian N dan T maka bentuk dari hasil akhir algoritma C4.5 yaitu pohon keputusan (decision tree) dapat langsung digambarkan seperti pada Gambar 2.



Gambar 2. Pohon Keputusan Hasil Akhir Paket Perawatan Wajah.

Pohon keputusan yang pertama diambil dari atribut Paket Perawatan Wajah yang merupakan atribut produk teratas pada Tabel 1. Contoh ini dipilih secara acak, bukan berdasarkan suatu perhitungan, kategori maupun kriteria tertentu. Berdasarkan pohon keputusannya dapat diketahui penjualan pada atribut produk tersebut mengalami penurunan sehingga dalam prediksi penjualan untuk masa mendatang dapat diperkirakan untuk melakukan persediaan barang tidak melebihi stok pada bulan terakhir yang dianalisis yaitu penjualan pada bulan Maret 2019. Dan untuk produk lain yang dibuatkan

pohon keputusannya adalah atribut produk Sabun Pembersih Wajah seperti yang tertera pada Gambar 3.



Gambar 3. Pohon Keputusan Hasil Akhir Sabun Pembersih Wajah.

Sementara pada pohon keputusan pada gambar 3 untuk atribut produknya memiliki nilai yang sempat mengalami peningkatan atau N dan T yang terjadi dari bulan Januari menuju bulan Februari. Hal ini dapat menjadi salah satu acuan yang mempengaruhi hasil prediksi penjualan pada masa mendatang karena terjadi fluktuasi atau turun naik angka penjualan. Pohon keputusan pada atribut Sabun Pembersih Wajah memiliki lebih banyak cabang dibandingkan dengan atribut Paket Perawatan Wajah seperti pada gambar 2. Hal tersebut terjadi karena perbedaan angka penjualan. Meskipun jumlahnya lebih rendah namun variasi kategori N dan T nya berbeda sehingga banyak kemungkinan melalui cabang-cabang pohon keputusannya.

Lalu untuk nilai hasil perhitungan entropy dan gain yang telah dilakukan, diketahui nilai entropy pada keseluruhan atribut partisi yaitu 1,014930619. Sementara itu untuk contoh perhitungan yaitu pada bulan Februari untuk kategori N adalah 0,528320834. Hal tersebut membuktikan bahwa atribut produk yang penjualannya naik masih lebih banyak dari pada atribut produk yang penjualannya mengalami penurunan meskipun perbedaan angkanya hanya nol koma sekian saja.

KESIMPULAN

Berdasarkan analisis dan perhitungan data penjualan pada Toko Online XXX pada bulan Januari hingga Maret 2019, didapatkan fakta bahwa lebih banyak produk yang penjualannya menurun dibanding yang meningkat. Hal tersebut diketahui dari banyaknya jumlah kategori N sebanyak 12 sedangkan kategori T sebanyak 19. Dari total 36 partisi, hanya 31 saja yang mengalami penjualan sedangkan 5 sisanya tidak ada penjualan sama sekali sehingga tidak dapat ditentukan untuk dikelompokkan pada kategori N maupun T. Sementara untuk produk dengan angka penjualan yang baik adalah Madu Rematik dan Madu Lambung yang mencapai angka ratusan per produknya pada setiap bulan. Namun pada dasarnya faktor yang mempengaruhi penjualan juga dapat ditentukan dari seberapa banyak dan seberapa seringnya setiap produk diiklankan agar menarik lebih banyak peminat sehingga penjualan dapat meningkat. Di sisi lain, penyediaan stok produk yang angka penjualannya kecil mungkin pada masa mendatang dapat dikurangi atau pemasaran produk melalui iklan dapat lebih ditingkatkan agar tidak ada stok produk yang terbuang karena permasalahan tenggat kadaluarsa atau produk yang rusak seiring waktu.

DAFTAR PUSTAKA

- Kemp, E., Childers, C. Y., and Williams, K. H, "Place branding: creating self-brand connections and brand advocacy, *Journal of Product & Brand Management*", 2012, Vol. 21, No.7, pp. 508–515.
- Kasapi, Irsi. and Cela, Ariana, "Destination Branding: A Review of the City Branding Literature", *Mediterranean Journal of Social Sciences*, 2017, Vol. 8 No 4 July.
- Wood L, "Brands and brand equity: definition and management, *Journal of Management Decision*", 2000, Vol.38, No. 9, pp.662 – 669.
- Knox, S. and Bickerton, D, "The six conventions of corporate branding, *European Journal of Marketing*", 2003, Vol. 37, No. 7/8, pp. 998 – 1016.
- Kavaratzis, M, "From city marketing to city branding: Towards a theoretical framework for developing city brands, *Journal of Place Branding*", 2004, Vol.1, No.1, pp.58 – 73.