

Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang *E-commerce*

G L Pritalia^{*1}

¹Departemen Teknik Elektro Teknologi Informasi, Fakultas Teknik, Universitas Gadjah Mada

E-mail: generosapritalia@gmail.com^{*1}

Masuk: 30 Agustus 2018, direvisi: 31 Agustus 2018, diterima: 31 Agustus 2018

Abstrak. Dalam sebuah perusahaan *e-commerce*, stok barang merupakan salah satu faktor penting yang harus diperhatikan oleh perusahaan. Jumlah barang masuk maupun keluar perlu diperhatikan, tujuannya untuk menjaga stok yang tersedia dalam gudang menjadi stabil. Stok stabil yang dimaksud dalam gudang adalah barang tidak kelebihan stok maupun barang tidak kehabisan stok. Permasalahan yang dihadapi pada penjualan barang adalah setiap barang memiliki minat beli yang berbeda-beda. Barang dengan minat beli yang tinggi akan cepat habis dibandingkan dengan barang yang memiliki minat beli rendah. Untuk itu, stok barang perlu ditambah pada waktu stok barang menipis. Algoritma C4.5 merupakan kelompok algoritma dengan menggunakan pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Semakin kaya informasi atau pengetahuan yang dikandung oleh data *training*, maka akurasi akan semakin meningkat. Algoritma ini digunakan untuk menganalisis waktu pembelian stok barang yang sudah menipis dengan mengklasifikasi barang mana yang sudah waktunya di tambah stok maupun belum, sehingga ketersediaan barang tetap stabil dan terjaga. Hasil dari analisis menggunakan algoritma C4.5 adalah untuk menentukan waktu penentuan ketersediaan barang memiliki tingkat keakuratan sebesar 98.9%.

Kata kunci: E-commerce; stok; ketersediaan barang; Algoritma C4.5; klasifikasi

Abstract. For an e-commerce company, goods stock belongs to one of many important aspects. A number of in and out stocks need an extra care due to keeping goods' stability in the warehouse. A stable stock means the goods are not too many or lacking, even out of stock. The most common problem is the difference of customers' interests toward each goods. Surely, customers' favorite goods will be out of stock faster than the ones that are not best seller. Therefore, a restocking must be actioned when they start lacking. Algorithm C4.5 is an algorithm group using Decision Tree. A decision tree is a strong and famous classification method and prediction. The richer one information or knowledge contained by data training is, the more accurate it will be. This algorithm is used to analyze the good's purchase time when it starts lacking. This will also classify the right time for restocking so the goods' availability will keep stable. From the analysis using C4.5 algorithm to define a perfect restocking time and goods' availability have an accuracy of 98.9%.

Keywords: E-commerce; stock; Availability of Goods; Algoritma C4.5; classification

1. Pendahuluan

Beberapa tahun terakhir tingkat bisnis melalui *e-commerce* semakin meningkat. Setiap *e-commerce* mempunyai proses bisnisnya masing-masing, sebagai contoh penjualan barang fisik maupun jasa, dan sebagainya. *E-commerce* yang bergerak dibidang jual beli barang fisik, tentunya membutuhkan gudang sebagai penyimpanan sementara barang yang hendak dijual.

Persediaan barang didalam gudang perlu dijaga setiap harinya. Jumlah penjualan yang fluktuatif mengakibatkan stok barang yang tersedia tidak stabil dan dapat berdampak langsung ke konsumen. Kepuasan konsumen dalam menggunakan *e-commerce* dapat berkurang karena tidak tersedianya barang yang hendak dibeli. Ketersediaan barang yang tidak dikelola dengan baik juga berdampak pada perusahaan, misalkan barang habis pada saat permintaan konsumen tinggi maka yang akan terjadi adalah permintaan barang harus diundur ataupun dibatalkan sehingga berdampak langsung ke penjualan *e-commerce*.

Permasalahan lainnya adalah penimbunan stok barang yang berlebihan. Kesalahan prediksi penjualan menjadi salah satu alasan yang membuat *e-commerce* membeli stok barang dalam jumlah besar yang akhirnya tidak habis dijual sehingga stok barang tersebut membengkak di gudang. Penimbunan ini menyebabkan *e-commerce* merugi dikarenakan dana yang harus keluar untuk proses penyimpanan barang. Selain itu, terdapat juga istilah *backorder*, dimana barang yang sudah dipesan konsumen belum tersedia, sehingga konsumen harus menunggu sampai barang tersebut tersedia.

Permasalahan - permasalahan yang terjadi diatas disebabkan karena kesulitan menentukan stok minimum tiap barang yang harus dipenuhi berdasarkan minat konsumen[1]. Penentuan stok barang dapat ditentukan dari berbagai aspek. Pada penelitian ini data yang akan digunakan untuk menentukan waktu pembelian stok barang adalah data *backorder*. Dari data tersebut akan dilihat faktor yang mempengaruhi *backorder* sehingga *e-commerce* dapat memperkirakan stok barang sebelum dipesan oleh konsumen.

Algoritma yang akan digunakan pada penelitian ini adalah algoritma C4.5. Algoritma ini akan menentukan faktor-faktor yang mengindikasi barang tersebut perlu di tambah stoknya. Berdasarkan faktor-faktor tersebut, diharapkan stok barang di dalam gudang menjadi stabil dan terhindar dari *backorder*.

2. Kerangka Teoritis

2.1. E-commerce

Electronic Commerce (E-commerce) adalah proses pembelian, penjualan atau pertukaran produk, jasa dan informasi melalui jaringan komputer [2]. Sedangkan menurut Laudon dalam [3], *e-commerce* diartikan sebagai penggunaan internet dan *web* untuk bertransaksi bisnis atau bisa juga didefinisikan sebagai suatu transaksi perdagangan secara digital antar organisasi atau individu. Persediaan merupakan simpanan material yang berupa bahan mentah, barang dalam proses, dan barang jadi. Fungsi di perlukannya persediaan adalah untuk mempermudah jalannya operasi perusahaan yang dilakukan secara berurut-urut untuk proses bisnis [4]. Dalam perusahaan *e-commerce* stok barang merupakan salah satu hal terpenting, baik stok barang yang ada pada gudang perusahaan *e-commerce* itu sendiri maupun dari produsen. Permasalahan yang sering terjadi terkait stok barang yaitu terjadinya *backorder product*. *Backorder product* terjadi saat pembeli sudah memesan barang kepada *e-commerce/dropshipper*, namun di sisi produsen (*e-commerce*) tiba-tiba stok barang telah habis karena telah diborong oleh konsumen yang telah membeli terlebih dahulu [5].

2.2. Data Mining

Penambangan data atau *data mining* berada di antara cabang ilmu komputer dan statistik, dengan memanfaatkan kemajuan dari kedua disiplin ilmu tersebut untuk mengekstraksi informasi dari besar dari *database* [6]. *Data mining* merupakan serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini belum diketahui dari suatu kumpulan data. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk

mengekstrasi dan mengidentifikasi informasi yang bermanfaat serta pengetahuan yang terkait dari berbagai *database* besar [7].

2.3. Algoritma C4.5 / Decision Tree

Algoritma C4.5 merupakan salah satu teknik klasifikasi pada *machine learning* yang digunakan pada proses *data mining* dengan membentuk sebuah pohon keputusan (*decision tree*) yang direpresentasikan dalam bentuk aturan [8]. Algoritma C4.5 merupakan kelompok algoritma dengan menggunakan pohon keputusan [9]. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal [10]. Semakin kaya informasi atau pengetahuan yang dikandung oleh data *training*, maka akurasi akan semakin meningkat. Pohon dalam analisis pemecahan masalah pengambilan keputusan adalah pemetaan mengenai alternatif-alternatif pemecahan masalah yang dapat diambil dari masalah tersebut. Pohon tersebut juga memperlihatkan faktor-faktor kemungkinan/probabilitas yang akan mempengaruhi alternatif-alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila kita mengambil alternatif keputusan tersebut. Pengambilan keputusan merupakan masalah penting bagi organisasi untuk menemukan alternatif terbaik dari alternatif yang ada [11].

C4.5 adalah algoritma yang dibuat oleh Ross Quinlan dan digunakan untuk membuat pohon keputusan. Algoritma ini sering dikategorikan sebagai pengklasifikasi statistik. C4.5 merupakan pengembangan dari algoritma ID3 yang menggunakan entropi informasi, atribut kontinyu dan diskret, atribut kategorial dan numerik, dan *missing values* [12].

Algoritma ini membutuhkan set data latih karena termasuk algoritma pembelajaran yang terawasi (*supervised learning algorithm*). Set data latih berupa sampel yang sudah terklasifikasi. C4.5 menganalisa set data latih dan membangun pengklasifikasian yang harus secara tepat mampu mengklasifikasikan data latih maupun data uji.

Formulasi Matematis dari algoritma C 4.5 dapat dilihat pada formula (1)-(5).

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

dengan penjelasan \log_2 :

$$\log_2(x) = \frac{\ln(x)}{\ln(2)} \quad (2)$$

dengan:

S = ruang (data) sampel yang digunakan untuk training.

p_i = proporsi dari S_i terhadap S. S_i diperoleh dari *Gain*.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{|S|} \frac{|S_i|}{|S|} * Entropy(S_i) \quad (3)$$

dengan :

S = ruang (data) sampel yang digunakan untuk training.

A = atribut.

V = suatu nilai yang mungkin untuk atribut A.

Nilai(A) = himpunan yang mungkin untuk atribut A.

$|S_i|$ = jumlah sampel untuk nilai i.

$|S|$ = jumlah seluruh sampel data.

$Entropy(S_i)$ = entropi untuk sampel-sampel yang memiliki nilai i.

$$SplitInfo(S, A) = - \sum_{i=1}^{|S|} \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (4)$$

dengan:

- S = ruang (data) sampel yang digunakan untuk training.
- A = atribut.
- S_i = jumlah sampel untuk atribut i

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)} \quad (5)$$

dengan:

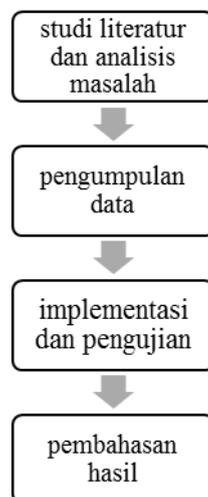
- S = ruang (data) sampel yang digunakan untuk training.
- A = atribut.
- Gain(S,A) = *information gain* pada atribut A
- SplitInfo(S,A) = *split information* pada atribut A

Secara umum langkah dalam membuat pohon keputusan menggunakan algoritma C4.5 adalah sebagai berikut:

1. Menghitung entropi total dari dataset dilanjutkan dengan entropi masing-masing atribut.
2. Setelah diperoleh entropi masing-masing atribut, menghitung *information gain* masing-masing.
3. Memilih atribut yang memiliki *information gain* paling besar sebagai akar.
4. Mengulangi perhitungan entropi dan *gain* untuk menentukan atribut berikutnya sebagai daun.

3. Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini terbagi dalam beberapa subtahap seperti ditunjukkan pada gambar 1.



Gambar 1. Diagram flow metodologi penelitian.

Pada tahap awal dimulai dengan melakukan identifikasi terhadap masalah yang muncul dan melakukan beberapa studi literatur, jurnal, artikel, situs internet, buku-buku, dan karya ilmiah. Tahap ini dilakukan untuk mendapatkan informasi terkait Sistem Rekomendasi dan Algoritma C4.5. Informasi didapatkan dari mengamati permasalahan yang berhubungan dengan faktor – faktor yang dibutuhkan untuk dapat merekomendasi produk *e-commerce* dan mengamati juga penelitian-penelitian terkait rekomendasi produk *e-commerce* yang mempengaruhi ketersediaan produk di perusahaan

dengan menggunakan metode C4.5 lainnya. Berdasarkan permasalahan – permasalahan yang ada, selanjutnya dianalisa untuk mengetahui ruang lingkup permasalahan yang akan diteliti.

Tahapan berikutnya adalah analisis kebutuhan data dan mengumpulkan data. Kegiatan ini dilakukan untuk mencari dan mengumpulkan data. Data akan analisa untuk menyesuaikan proses data yang akan diuji dengan metode C4.5. Data penelitian ini diambil dari portal data Kaggle [13]. Data penelitian merupakan dataset yang berupa produk retail *backorder*. Data awal penelitian ini memiliki jumlah 5.000 data dengan 23 atribut. Data mentah tersebut masih berisi banyak data yang tidak penting dan atribut yang tidak mendukung. Oleh karena itu diperlukan proses pengolahan data awal untuk memperoleh data yang berkualitas.

Setelah menjalankan tahap analisis kebutuhan data dan mengumpulkan data maka selanjutnya akan dilakukan implementasi dan pengujian, sesuai dengan pengolahan data maka pada tahap implementasi adalah data akan diolah dengan menggunakan *tools*. *Tools* yang digunakan dalam penelitian ini adalah *software* Weka. Algoritma C4.5 merupakan pengembangan dari ID3, sedangkan pada perangkat lunak *open source* WEKA mempunyai versi sendiri C4.5 yang dikenal sebagai J48. Pengujian dilakukan untuk mencari validitas dengan membandingkan aturan klasifikasi yang dihasilkan algoritma C4.5 oleh WEKA. Pada tahapan ini terbagi lagi menjadi beberapa sub bagian, diantaranya pengolahan data awal dan pengujian metode.

3.1. Pengolahan Data Awal

Data yang masih mengandung data-data yang tidak penting atau kosong/hilang mungkin menyebabkan data tidak valid atau bahkan tidak dapat diproses, maka dari itu untuk mendapatkan data yang valid dan berkualitas, terdapat beberapa teknik yang harus dilakukan seperti pembersihan dan transformasi data.

3.1.1. Pembersihan Data

Proses pembersihan data yaitu proses dimana data-data yang yang tidak penting atau tidak lengkap akan dihapus[14]. Proses selanjutnya adalah mengisi data-data yang kosong/hilang, data yang tidak lengkap.

3.1.2 Transformasi Data

Transformasi data ini bertujuan untuk meningkatkan akurasi dan efisiensi algoritma[15]. Algoritma C4.5 dapat mengelolah data numerik (kontinyu) dan diskret sehingga nilai-nilai dari setiap atribut yang terdapat pada data set tidak perlu ditransformasikan karena data set sudah sesuai dengan syarat[16].

3.2. Pengujian Metode

Pengujian penelitian ini dilakukan menggunakan algoritma C4.5 (*decision tree*) untuk menilai produk *retail* yang akan mengalami *backorder*. Setelah melalui tahap pengolahan data awal dan menjadikan data tersebut valid maka dataset tersebut telah siap untuk dijalankan pada aplikasi Weka versi 3.9. Tahapan terakhir adalah penarikan kesimpulan dari hasil yang sudah dianalisis. Hasil tersebut nantinya dapat digunakan untuk membuat keputusan yang berkaitan dengan pembelian stok barang maupun laporan untuk jumlah minimum barang yang tersedia.

4. Hasil dan Pembahasan

4.1. Data Penelitian

Data penelitian merupakan dataset yang berupa produk retail *backorder*. Data awal penelitian ini memiliki jumlah 5.000 data dengan 23 atribut. Data mentah tersebut masih berisi banyak data yang tidak penting dan atribut yang tidak mendukung. Oleh karena itu agar mendapatkan data yang berkualitas dalam Data Mining maka diperlukan proses pengolahan data awal. Berikut ini merupakan data mentah dari produk retail *backorder*.

Tabel 1. Tabel Data Set Penelitian

Sku	Nat_Inv	Fore_3_Month	Sales_1_Month	Pieces_past_due	BackOrder
3285085	62	0	0	0	“No”
3290902	1169	922	3483	0	“No”
3352105	0	0	0	0	“Yes”
3285358	17	0	0	0	“No”

Karena data memiliki jumlah 5.000 data dengan 23 atribut yang terlalu banyak maka hanya menggunakan beberapa sampel seperti yang ditunjukkan pada tabel 1. Berikut ini merupakan penjelasan dari 23 atribut yang akan digunakan:

1. Sku merupakan ID dari produk.
2. National_Inv merupakan tingkat persediaan saat ini untuk bagian tersebut.
3. Lead_Time merupakan waktu transit untuk produk (jika tersedia).
4. In_Transit_Qty merupakan jumlah produk dalam perjalanan dari sumber.
5. Forecast_3_Month merupakan penjualan perkiraan untuk tiga bulan ke depan.
6. Forecast_6_Month merupakan penjualan perkiraan untuk enam bulan ke depan.
7. Forecast_9_Month merupakan penjualan perkiraan untuk sembilan bulan ke depan.
8. Sales_1_Month merupakan kuantitas penjualan untuk periode waktu satu bulan sebelumnya.
9. Sales_3_Month merupakan kuantitas penjualan untuk periode waktu tiga bulan sebelumnya.
10. Sales_6_Month merupakan kuantitas penjualan untuk periode waktu enam bulan sebelumnya.
11. Sales_9_Month merupakan kuantitas penjualan untuk periode waktu sembilan bulan sebelumnya.
12. Min_Bank merupakan jumlah minimal untuk persediaan.
13. Potential_Issue merupakan isu sumber untuk bagian yang teridentifikasi.
14. Piece_Past_Due merupakan bagian jatuh tempo dari sumber.
15. Perf_6_Month_Avg merupakan sumber kinerja untuk periode enam bulan sebelumnya.
16. Perf_12_Month_Avg merupakan sumber kinerja untuk periode 12 bulan sebelumnya.
17. Local_Bo_Qty merupakan jumlah pesanan saham terlambat.
18. Deck_Risk merupakan bagian produk risiko.
19. Oe_Constraint merupakan bagian produk risiko.
20. Ppap_Risk merupakan bagian produk risiko.
21. Stop_Auto_Buy merupakan bagian produk risiko.
22. Rev_Stop merupakan bagian produk risiko.
23. Went_On_Backorder merupakan produk sebenarnya di *backorder*. Ini adalah nilai target.

4.2. Pengolahan Data Awal

Data yang masih mengandung data-data yang tidak penting, kosong, atau hilang mungkin menyebabkan data tidak valid atau bahkan tidak dapat di proses, maka dari itu untuk mendapatkan data yang valid dan berkualitas perlu dilakukan pembersihan data.

4.2.1 Pembersihan Data

Tabel 2 menunjukkan hasil dari data mentah yang telah mengalami proses pembersihan data.

Tabel 2. Tabel hasil pembersihan data

Sku	Nat	Fore_3	Sales_1	Piece	BackOrder
3285085	62	0	0	0	No
3290902	1169	922	3483	0	No
3352105	0	0	0	0	Yes

Sku	Nat	Fore_3	Sales_1	Piece	BackOrder
3285358	17	0	0	0	No

Dari hasil pada tabel 2 dapat dipastikan bahwa proses pembersihan data di mana data yang tidak penting yaitu tanda “ (petik dua) akan dihapus, selanjutnya untuk data yang masih bernilai *null* akan diberikan nilai 0 sehingga data tersebut menjadi lengkap dan siap untuk diproses ke tahap selanjutnya.

4.2.2 Transformasi Data

Nilai-nilai dari setiap atribut yang terdapat pada dataset tidak perlu ditransformasikan karena dataset sudah sesuai dengan syarat. Pada dataset terdapat sejumlah 16 atribut data yang bertipe integer yaitu sku, nat_inv, lead_time, in_transit_qty, forecast_3_month, forecast_6_month, forecast_9_month, sales_1_month, sales_3_month, sales_6_month, sales_9_month, min_bank, pieces_past_due, perf_6_month_avg, perf_12_month_avg, local_bo_qty dan 7 atribut data yang bertipe numerik yaitu potential_issue, deck_risk, oe_constraint, pppap_risk, stop_auto_buy, rev_stop, dan went_on_backorder.

4.3. Pengujian Metode

Pengujian penelitian ini dilakukan menggunakan Algoritma C4.5 (*decision tree*) untuk menilai produk *retail* yang akan mengalami *backorder*. Setelah melalui tahap pengolahan data awal dan menjadikan data tersebut valid maka data set tersebut telah siap untuk dijalankan pada aplikasi Weka versi 3.9. Hasil *Stratified Cross-Validation* ditampilkan pada tabel 3. Dari tabel 3 dapat terlihat bahwa tingkat kebenaran dari hasil klasifikasi berjumlah 4.948 dari 5.000 data dengan keakuratan kebenaran sebanyak 98,9798%. Tingkat kesalahan dari hasil klasifikasi berjumlah 51 dari 5000 data dengan keakuratan kesalahan sebanyak 1,0202%. Statistik kappa bernilai 0.1875 serta kesalahan yang absolut bernilai 0.0179. Kesalahan dari akar bernilai 0.0992, kesalahan absolut relatif sebesar 85.8504%, dan kesalahan dari akar relatif sebanyak 97.8122% sehingga jumlah *instance* yang dieksekusi sebanyak 4.999 data dalam waktu 0,38 detik.

Tabel 3. Tabel Hasil Stratified cross-validation

Correctly Classified Instances	4948	98.9798%
Incorrectly Classified Instances	51	1.0202%
Kappa statistic		0.1875
Mean absolute error		0.0179
Root mean squared error		0.0992
Relative absolute error		85.8504%
Root relative squared error		97.8122%
Total Number of Instances	4999	

Tabel 4. Tabel *Detailed Accuracy By Class*

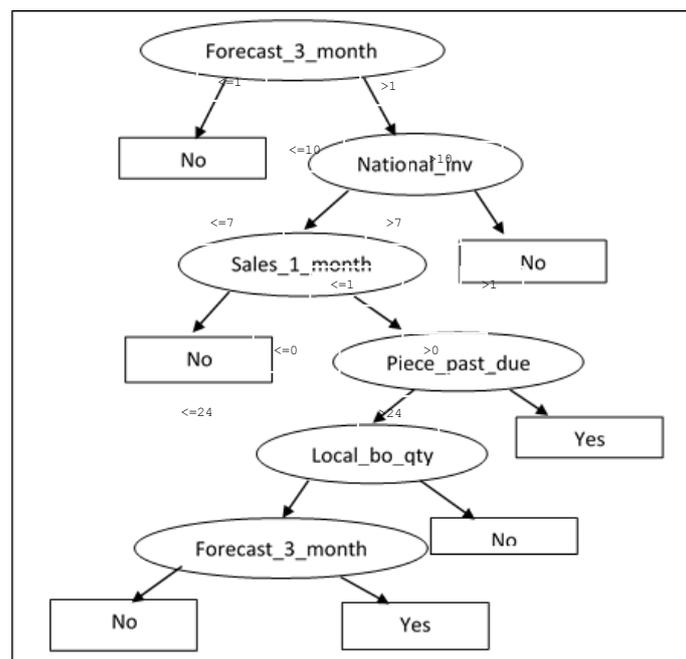
TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.999	0.855	0.991	0.999	0.995	0.248	0.773	0.995	No
0.115	0.001	0.545	0.115	0.190	0.248	0.773	0.110	Yes
0.990	0.875	0.986	0.990	0.986	0.248	0.773	0.986	

Pada tabel 3 juga diketahui bahwa tingkat akurasi mencapai lebih dari 98% (terlihat pada bagian *Correctly Classified Instances*) Hal ini menunjukkan bahwa dalam pengolahan data awal terbukti bahwa tingkat akurasi tinggi dan dijalankan dalam waktu yang cepat atau dengan kata lain dapat dikatakan bahwa data tersebut adalah valid.

Keakuratan tersebut dapat didukung dengan hasil yang ditunjukkan pada tabel 4 yaitu dari hasil *Precision* dan *recall*. *Precision* dapat diartikan sebagai ketepatan atau kecocokan yang merupakan

perbandingan jumlah dokumen yang relevan terhadap *query* dengan jumlah dokumen yang diambil dari hasil pencarian. Pada penelitian ini diperoleh nilai *precision* tinggi yaitu 0.991 di kelas No dan 0.545 di kelas Yes. Sedangkan *recall* merupakan perbandingan jumlah dokumen relevan yang diambil sesuai dengan *query* yang diberikan dengan total kumpulan dokumen yang relevan dengan *query* yang bernilai 0.999 di kelas No dan 0.190 di kelas Yes. Kemudian dari hasil *recall* dan *precision* tersebut dapat menentukan *f-measure* yang merupakan suatu perhitungan evaluasi dalam temu kembali informasi yang mengkombinasikan *recall* dan *precision* dengan nilai 0.995 pada kelas No dan nilai 0.190 pada kelas Yes.

Dengan menggunakan algoritma C4.5 dapat diketahui informasi lain berupa pohon keputusan yang dapat menentukan apakah produk *retail* tersebut akan mengalami *backorder* atau tidak. Hasil struktur pohon sesuai dengan data set yang telah digunakan ditunjukkan pada gambar 2.



Gambar 2. Hasil Pohon Keputusan.

Keterangan gambar 2:

1. forecast_3_month : penjualan perkiraan untuk tiga bulan ke depan.
2. sales_1_month : kuantitas penjualan untuk periode waktu satu bulan sebelumnya.
3. national_inv : persediaan terakhir saat ini.
4. pieces_past_due : bagian jatuh tempo dari sumber.
5. local_bo_qty : jumlah pesanan saham terlambat

Dari gambar 2 dapat terlihat bahwa:

1. Jika Penjualan dengan perkiraan untuk tiga bulan ke depan bila kurang dari -1 maka masuk ke dalam tahap ketersediaan barang dalam posisi tersedia
2. Jika nilai lebih dari satu maka masuk ke dalam tahap persediaan terakhir saat ini
3. Jika kurang dari sama dengan 10 maka masuk ke dalam tahap bagian kuantitas penjualan untuk periode waktu satu bulan sebelumnya.
4. Jika lebih dari 10 maka ketersediaan barang dalam posisi tersedia.
5. Jika kuantitas penjualan untuk periode waktu satu bulan sebelumnya kurang dari sama dengan tujuh maka ketersediaan barang dalam posisi tersedia.
6. Jika lebih dari tujuh maka masuk ke dalam tahap jatuh tempo dari sumber.

7. Jika kurang dari sama dengan satu maka masuk ke dalam tahap jumlah pesanan saham terlambat.
8. Jika lebih dari satu maka ketersediaan barang dalam posisi tidak tersedia.
9. Jika pada tahap jumlah pesanan saham terlambat kurang dari sama dengan nol maka masuk ke dalam tahap penjualan perkiraan untuk tiga bulan ke depan.
10. Jika lebih dari nol maka ketersediaan barang dalam posisi tidak tersedia.
11. Jika pada tahap penjualan perkiraan untuk tiga bulan ke depan kurang dari sama dengan 24 maka ketersediaan barang dalam posisi tersedia.
12. Jika lebih dari 24 maka ketersediaan barang dalam posisi tidak tersedia.

5. Kesimpulan

Berdasarkan hasil implementasi dan uji coba aplikasi, dapat disimpulkan bahwa:

1. Data mining dengan algoritma C4.5 dapat diimplementasikan untuk penentuan ketersediaan barang *e-commerce* dengan dua kategori yaitu barang tersebut dalam kondisi tidak atau belum tersedia untuk pemesanan barang untuk pelanggan/*backorder* dan barang dalam kondisi tersedia untuk pemesanan barang untuk pelanggan.
2. Aplikasi Weka berhasil menentukan ketersediaan barang *e-commerce* dengan presentase keakuratan sebesar 98% dari 5000 sampel data dan 23 atribut dengan kecepatan waktu menjalankan program selama 0.38 detik.
3. Hasil penentuan ketersediaan barang *e-commerce* dari aplikasi penelitian ini dapat membantu bagian perusahaan untuk mengetahui status barang yang siap untuk diperjualbelikan. Hal ini dapat menjadi rekomendasi pengambilan keputusan dalam menerima pesanan agar barang tersebut tetap terjaga ketersediaannya dan tidak mengalami *backorder*. Hal ini penting agar pelanggan tetap terjaga kepercayaannya dalam melakukan transaksi barang.

6. Referensi

- [1] E. Muningsih and S. Kiswati, "Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang," *J. Bianglala Inform.*, 2015.
- [2] D. Irmawati, "Pemanfaatan E-Commerce Dalam Dunia Bisnis," *J. Ilm. Orasi Bisnis – ISSN 2085-1375 Ed. Ke-VI, Novemb. 2011*, 2011.
- [3] K. P. A. . Jonathan Sarwono, *Perdagangan Online: Cara Bisnis di Internet*. Jakarta: PT. Elex Media Komputindo, 2012.
- [4] A. Wijaya, M. Arifin, and T. Subiyanto, "Rancang Bangun Sistem Informasi Perencanaan Persediaan Barang," *JSIKA (Jurnal Sist. Inf. Komput. Akuntansi)*, vol. 2, no. 1, 2013.
- [5] A. Dyah Utami and R. Agus Triyono, "Pemanfaatan Blackberry Sebagai Sarana Komunikasi Dan Penjualan Batik Online Dengan Sistem Dropship Di Batik Solo 85," *Sentra Penelit. Eng. dan edukasi*, 2011.
- [6] E. Turban, J. E. Aronson, and T.-P. Liang, "Decision Support Systems and Intelligent Systems," *Decis. Support Syst. Intell. Syst.*, 2007.
- [7] D. K. Pane, "Implementasi Data Mining Pada Penjualan Produk Elektronik Dengan Algoritma Apriori (Studi Kasus : Kreditplus)," *Pelita Inform. Budi Darma*, 2013.
- [8] R. Soewono, R. Gernowo, and P. S. Sasongko, "Sistem Pakar Identifikasi Modalitas Belajar Siswa Dengan Implementasi Algoritma C4.5," *J. Sist. Inf. Bisnis*, vol. 4, no. 1, 2014.
- [9] B. D. Meilani and A. F. Slamet, "Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree," *J. IPTEK*, vol. 16, no. 1, pp. 1–6, 2012.
- [10] K. Hastuti, "Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif," *Semin. Nas. Teknol. Inf. Komun. Terap.*, 2012.
- [11] A. Darmawan, "Implementasi Simple Additive Weighting Untuk Monitoring Aktivitas Perkuliahan Dengan Menggunakan Radio Frequency Identification," *J. Sist. Inf. Bisnis*, vol. 7, no. 1, 2017.
- [12] L. Swastina, "Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa," *Gema Aktual.*, 2013.

- [13] "No Title." [Online]. Available: <https://www.kaggle.com/tiredgeek/predict-bo-trial>.
- [14] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Eeccis*, 2013.
- [15] M. Mirqotussa'adah, M. A. Muslim, E. Sugiharti, B. Prasetyo, and S. Alimah, "Penerapan Dizcretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4.5 dalam Mendiagnosa Diabetes," *Lontar Komput. J. Ilm. Teknol. Inf.*, 2017.
- [16] N. Jayanti, S. Puspitodjati, and T. Elida, "Teknik Klasifikasi Pohon Keputusan Untuk Memprediksi Kebangkrutan Bank Berdasarkan Rasio Keuangan Bank," in *Proceeding Seminar Ilmiah Nasional Komputer dan Sistem Intelijen (KOMMIT)*, 2008.