



Klasifikasi Produk Retur dengan Algoritma Pohon Keputusan C4.5

Amalia Anjani Arifyanti¹, Rhendy May Pradana², Indra Fajar Novian³

Jurusan Sistem Informasi, Fakultas Teknologi Informasi - ITATS^{1, 2, 3}

INFORMASI ARTIKEL

Jurnal IPTEK – Volume 22
Nomer 1, Mei 2018

ISSN:1411-7010
e-ISSN:2477-507X

Halaman:
79 – 86

DOI:
<https://dx.doi.org/10.31284/j.ipitek.2018.v22i1.243>

EMAIL

anjani.arifyanti@itats.ac.id

RIWAYAT ARTIKEL

Tanggal diterima :
29 April 2018

Tanggal diterbit :
20 Mei 2018

PENERBIT

LPPM- Institut Teknologi
Adhi Tama Surabaya

Jurnal IPTEK by LPPM-ITATS is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

ABSTRACT

Product returns always happen and it cannot avoided by company as seller of the product.. This process does not seem to be detrimental at first, but processing the information about product return can provide information that is beneficial for using it in the evaluation process. One way of processing this information is to classify the type of product returned and the reason. In this research, product that returned classify into three classes (strong, moderate, weak). The result can be use in evaluation in production process, packaging, and shipping. C4.5 decision tree algorithm is used because this algorithm best suitable for multi-classes classification, has good performance, and easily interpretable. Evaluation method used confusion matrix and cross-validation with 10 folds. The performance of classifier had 95,6%, 0,943, and 0,956 for accuracy, precision, and recall respectively.

Keywords: C4.5, Classification, Decision Tree, Product Returns

ABSTRAK

Pengembalian produk oleh konsumen kepada penjual bukan hal yang dapat dihindari. Aktivitas ini tidak terlihat merugikan pada awalnya, namun pengelolaan informasi dari pengembalian produk dapat memberikan informasi yang bermanfaat bagi penjual untuk digunakan dalam proses evaluasi. Salah satu cara pengelolaan informasi ini adalah dengan melakukan klasifikasi dari jenis produk yang dikembalikan oleh konsumen dan alasan-alasan yang menjadikan produk tersebut dikembalikan. Pada penelitian ini produk retur dimasukkan dalam tiga klasifikasi yaitu kuat, sedang, dan lemah. Hasil klasifikasi tersebut dapat menjadi evaluasi pada proses produksi hingga pengemasan dan pengiriman produk. Algoritma pohon keputusan C4.5 dipilih dalam penelitian ini karena algoritma ini dapat menangani multi-class, memiliki performa yang baik dalam klasifikasi, dan model klasifikasi yang dihasilkannya mudah dipahami. Pembagian data latih dan data uji dilakukan dengan menggunakan metode cross-validation 10 folds karena jumlah data yang terbatas. Model klasifikasi yang dihasilkan pada penelitian ini memiliki performa akurasi sebesar 95,6%, precision sebesar 0,943, dan recall sebesar 0,956.

Kata Kunci: C4.5, Klasifikasi, Pohon Keputusan, Produk Retur

PENDAHULUAN

Salah satu tantangan bagi para pelaku bisnis di bidang manufaktur adalah adanya pengembalian produk yang dilakukan oleh konsumen. Pengembalian produk terjadi karena konsumen tidak merasa puas terhadap produk yang telah dibeli. Ketidakpuasan ini muncul dikarenakan berbagai alasan misalnya produk yang dibeli cacat, tidak sesuai harapan, dan alasan lainnya. Barang yang telah dikembalikan tersebut dapat digantikan dengan produk yang baru (*exchange*), pengembalian uang (*refund*), dan diberikannya jasa tambahan [1]. Pengembalian produk merupakan salah satu aktivitas yang bernilai tambah rendah bagi

perusahaan. Aktivitas ini dapat mengurangi daya saing dan keuntungan perusahaan [2]. Aktivitas pengembalian produk ini tidak dapat dihilangkan dalam proses bisnis perusahaan. Hal ini dikarenakan pengembalian produk merupakan hak konsumen yang telah tercantum pada Undang-Undang Perlindungan Konsumen [3]. Walaupun terlihat merugikan pada awalnya, namun jika perusahaan jeli mengelola informasi mengenai produk apa saja yang dikembalikan konsumen dan alasan pengembaliannya, maka hal tersebut dapat menjadi dasar evaluasi bagi perusahaan [4]. Pengembalian produk tidak serta merta dapat dilakukan begitu saja oleh konsumen. Barang yang dapat dikembalikan oleh konsumen harus mengikuti aturan pengembalian (*return policy*) yang ditetapkan oleh perusahaan. Aturan pengembalian dinilai sangat penting karena pelanggan akan lebih memilih perusahaan yang memiliki aturan pengembalian yang jelas sehingga pelanggan tidak merasa dirugikan dan hal ini dapat memberikan perasaan aman bagi para pelanggan [5]. Adanya aturan pengembalian produk yang telah dibeli karena berbagai alasan dapat menjadi salah satu faktor dalam peningkatan niat beli konsumen [6].

Informasi mengenai produk apa saja yang dikembalikan oleh konsumen dan alasan-alasan pengembaliannya dapat menjadi dasar evaluasi bagi perusahaan. Dalam penelitian ini akan dilakukan klasifikasi produk-produk yang dikembalikan oleh konsumen beserta alasan pengembaliannya untuk memprediksi pengembalian produk di masa mendatang. Klasifikasi akan dilakukan dengan menggunakan algoritma pohon keputusan C4.5. Algoritma ini dinilai dapat menangani klasifikasi multi-class dengan baik [7]. Algoritma ini juga merupakan algoritma yang banyak digunakan dalam berbagai bidang penelitian karena model klasifikasi yang dihasilkannya mudah dipahami dan memiliki performa yang lebih baik dibandingkan algoritma pohon keputusan lainnya yaitu ID3 dan C5.0 [8]. Kasus pengembalian produk akan dikelompokkan ke dalam tiga kelas yaitu kuat, sedang, dan lemah yang merujuk pada banyaknya kasus pengembalian produk sejenis. Dengan adanya klasifikasi ini, diharapkan informasi ini menjadi evaluasi bagi tim produksi dalam proses produksi produk, bagi tim *Research and Development* (RnD) dalam pengembangan produk baru, hingga tim pengemasan dan pengiriman produk. Evaluasi ini diharapkan dapat meningkatkan kualitas produk perusahaan sehingga untuk ke depannya dapat meminimalisir jumlah produk yang dikembalikan oleh konsumen.

TINJAUAN PUSTAKA

Pengembalian Barang

Pengembalian produk (*returned goods*) merupakan aktivitas bernilai tambah rendah bagi perusahaan maupun pelaku bisnis [2]. Pengembalian produk terjadi jika produk yang diterima oleh konsumen tidak sesuai dengan keinginan konsumen atau produk yang diterima konsumen dalam keadaan rusak. Pengembalian produk ini terbagi menjadi dua jenis yaitu:

1. Pengembalian pembelian/ Retur pembelian
Pengembalian ini dilakukan oleh konsumen. Hal ini disebabkan adanya kerusakan produk yang diterima oleh konsumen atau produk yang diterima konsumen tidak sesuai dengan pesanan [9]. Kesalahan atau kerusakan produk bukan dikarenakan oleh konsumen namun karena kesalahan yang terjadi di sisi penjual.
2. Pengembalian penjualan/ Retur penjualan
Pengembalian ini terdapat pada sisi penjual. Aktivitas ini merupakan penerimaan produk yang dikembalikan oleh konsumen kepada penjual disebabkan produk yang diterima oleh konsumen tidak sesuai harapan konsumen [10].

Aktivitas pengembalian ini tidak dapat diabaikan oleh perusahaan dan para pelaku bisnis. Hal ini telah tercantum pada Undang-Undang Perlindungan Konsumen (UUPK) nomor 8 tahun 1999 pasal 4 mengenai hak konsumen [3]. Menurut undang-undang tersebut konsumen atau konsumen berhak mendapatkan produk sesuai dengan yang dijanjikan dengan dan jika tidak maka konsumen berhak mendapatkan kompensasi, ganti rugi dan/atau penggantian. Oleh karenanya, pelaku bisnis saat ini membuat aturan-aturan pengembalian produk yang harus ditaati oleh para konsumennya misalnya produk yang

dikembalikan harus dalam rentang waktu tertentu, kerusakan bukan diakibatkan pelanggan, dan sebagainya. Aturan pengembalian yang jelas menunjukkan bahwa pelaku bisnis memiliki kualitas yang baik.

Data Mining

Data mining atau penggalian data merupakan bagian dari *knowledge discovery in database* (KDD). Proses tersebut mencakup dari pengumpulan data sebagai masukan, pemrosesan data, hingga keluaran atau hasil olahan data. Secara umum proses pengolahan data hingga menjadi informasi dapat dilihat pada gambar 1 berikut ini.



Gambar 1. Proses KDD [11]

Masukan dari proses KDD adalah data mentah yang didalamnya terdapat informasi penting atau dibutuhkan oleh penggali informasi. Diperlukan proses pengolahan data mentah sebelum masuk dalam proses data mining. Tahap pengolahan data mentah ini berada dalam tahap pra-proses. Tahap ini digunakan untuk mengubah data menjadi format yang ditentukan dan dibutuhkan dalam proses data mining. Tahap ini juga termasuk dalam penggabungan data dari berbagai sumber menjadi satu, penghapusan *noise* dan data duplikat, dan pemilihan atribut yang dibutuhkan dan relevan. Tahap ini merupakan tahap yang menghabiskan banyak sumber daya, tetapi tahap ini merupakan tahapan penting dalam keseluruhan proses KDD. Hal ini disebabkan kualitas data masukan memberikan pengaruh terhadap hasil keluaran proses penggalian data. Tahap data mining dilaksanakan setelah tahap pra-proses data. Penggalian data dilakukan dengan observasi pola dan hubungan dalam data menggunakan algoritma *machine learning* [12]. Pola dan hubungan ini akan menghasilkan informasi yang berguna. Tahap terakhir adalah *post-processing*, dimana dalam tahap ini hasil penggalian data akan diintegrasikan dengan sistem pengambilan keputusan atau divisualisasikan dalam bentuk memudahkan pengguna dalam pembacaan informasi.

Data mining secara umum terbagi menjadi dua jenis [11] yaitu sebagai berikut:

1. Prediktif. Tujuan metode dalam jenis ini adalah untuk memprediksi nilai dari target atribut berdasarkan nilai dari atribut-atribut lainnya. Klasifikasi dan regresi termasuk ke dalam jenis prediktif.
2. Deskriptif. Tujuan dari metode data mining ini adalah untuk mendapatkan pola yang menghasilkan simpulan hubungan dari data. Metode ini digunakan untuk menjelaskan pola dan hubungan data. Clustering termasuk dalam metode deskriptif.

Klasifikasi

Klasifikasi merupakan salah satu *task* dalam data mining dan termasuk ke dalam jenis *supervised methods*. *Supervised methods* membutuhkan 1) Target atribut yang telah didefinisikan, dan 2) Data latih, pada dataset ini target atribut telah didefinisikan sehingga algoritma pembelajar dapat mempelajari pola atau hubungan antar atribut prediktor (*independent attribute*) terhadap atribut target (*dependent attribute*). Oleh karena klasifikasi merupakan *supervised methods* maka klasifikasi melalui dua tahap proses [13] yaitu 1) Tahap pembelajaran, dimana model klasifikasi dibuat, dan 2) Tahap klasifikasi, dimana model klasifikasi digunakan untuk memprediksi kelas, sebagai target atribut, dari *instance* yang belum diketahui kelasnya. Pada tahap pertama yaitu tahap pembelajaran, algoritma *machine learning* digunakan untuk menganalisis pola atau hubungan dari atribut data dalam menentukan kelas atau target atribut. Hasil analisis ini dibentuk ke dalam model klasifikasi. Model klasifikasi ini diuji dan dinilai performanya sebagai *classifier*. Jika model klasifikasi dinilai memiliki performa yang baik, maka model klasifikasi dapat diterapkan pada kasus nyata.

Algoritma pembelajar atau *machine learning* yang digunakan dalam klasifikasi memiliki jenis yang beragam. Beberapa contoh algoritma pembelajar yaitu k-nearest neighbor yang berdasarkan jarak antar atribut pada masing-masing instance, pohon keputusan (*decision tree*) yang model klasifikasinya dibentuk ke dalam bentuk pohon, naive bayes yang berdasarkan kemungkinan/probability, neural network, dan algoritma lainnya. Pemilihan algoritma pembelajar dipengaruhi oleh ukuran, kualitas (*noise, missing value*, dan sebagainya), dan karakteristik dari data yang digunakan dalam klasifikasi [14]. Algoritma pembelajar akan mempengaruhi kecepatan dalam proses pembelajaran (*training time*) dan performa (akurasi dalam prediksi) dari model klasifikasi yang dihasilkan.

Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma pohon keputusan yang banyak digunakan pada penelitian mengenai klasifikasi [15]. C4.5 memiliki performa yang cukup baik jika dibandingkan dengan algoritma pohon keputusan lainnya [8] dan algoritma pembelajar lainnya [16]. C4.5 memiliki kemampuan yang baik dalam hal kecepatan pembelajaran data latih dan tingkat error yang cukup rendah. Seperti algoritma pohon keputusan lainnya, algoritma C4.5 membentuk model klasifikasi dalam bentuk struktur pohon. Dalam struktur tersebut terdapat 1) *Internal-node* yang merepresentasikan atribut, dengan node paling atas disebut dengan *root*, 2) *Branch* atau cabang yang merepresentasikan luaran dari atribut, dan 3) *Leaf-node* yang merupakan target atribut atau kelas [13]. Secara umum tahapan pembentukan pohon keputusan menggunakan C4.5 adalah sebagai berikut [11]:

1. Menyiapkan data latih. Data latih diambil dari data terdahulu yang telah memiliki kelas sehingga dapat menjadi acuan bagi prediksi kelas pada data uji.
2. Pilih atribut sebagai *node*. Pemilihan atribut ini ditentukan melalui penghitungan nilai gain dari masing-masing atribut. Atribut yang memiliki nilai Gain paling tinggi akan menjadi *node*. Berikut adalah rumus Gain:

$$\Delta = I(\text{parent}) - \sum_{j=1}^N \frac{N(v_j)}{N} I(v_j) \dots\dots\dots (1)$$

Dengan:

$I(\text{parent})$ = nilai ketidak murnian *parent* node (direpresentasikan dengan nilai entropy dari *parent* node)

$I(V_j)$ = nilai ketidakmurnian *child* node

N = jumlah *record* dalam *parent* node

$N(V_j)$ = jumlah *record* dalam *child* node

Nilai entropy dihitung terlebih dahulu untuk mendapatkan nilai gain, nilai entropy dihitung dengan rumus sebagai berikut:

$$\text{Entropy}(t) = -\sum_{i=0}^{c-1} p(i|t) \log_2 p(i|t) \dots\dots\dots (2)$$

Dengan:

p = pecahan

i = jumlah *record* pada kelas i

t = jumlah *record* dalam node

3. Buat cabang untuk setiap *node*.
4. Ulangi tahap 2 dan 3 hingga:
 - a. Seluruh *record* pada *node* masuk dalam kelas yang sama atau entropy bernilai 0
 - b. Seluruh atribut telah digunakan
 - c. Node kosong atau tidak memiliki *record*

Evaluasi

Sebelum model klasifikasi diaplikasikan dalam kasus nyata, pengklasifikasi tersebut harus melalui proses evaluasi untuk mengukur tingkat performa model klasifikasi yang dihasilkan. Pengukuran performa model klasifikasi dapat melalui beberapa jenis model evaluasi yang salah satunya adalah *confusion matrix*. *Confusion matrix* dasar berdimensi 2x2 yang terdiri dari baris horisontal adalah kelas prediksi (*predicted*

classes) dan kolom vertikal adalah kelas sebenarnya (*actual classes*). Kelas yang digunakan dalam matrix dasar terbagi menjadi dua yaitu kelas Yes dan No. *Confusion matrix* dasar dapat dilihat pada tabel 1.

Tabel 1. *Confusion Matrix* [17]

<i>Predicted Class</i>	<i>Actual Class</i>		
		Yes	No
	Yes	TP (<i>True Positive</i>)	FP (<i>False Positive</i>)
No	FN (<i>False Negative</i>)	TN (<i>True Negative</i>)	

Keterangan:

- TP = kelas prediksi adalah Y namun kelas sebenarnya adalah Y. *Correct result*.
- FP = kelas prediksi adalah Y namun kelas sebenarnya adalah N. *Unexpected result*.
- FN = kelas prediksi adalah N namun kelas sebenarnya adalah Y. *Missing result*.
- TN = kelas prediksi adalah N namun kelas sebenarnya adalah N. *Correct absence of result*.

Dari *confusion matrix* tersebut, dihitung akurasi, precision, dan recall untuk mengetahui performa dari model klasifikasi.

1. Akurasi

Akurasi merupakan kemampuan pengklasifikasi dalam memprediksi kelas dari masing-masing instance dalam kelas yang benar.

$$\frac{(TP + TN)}{(TP + FP + TN + FN)} \dots\dots\dots (1)$$

2. Precision

Precision merupakan kemampuan pengklasifikasi dalam memprediksi kelas yang benar-benar relevan.

$$\frac{TP}{(TP + FP)} \dots\dots\dots (2)$$

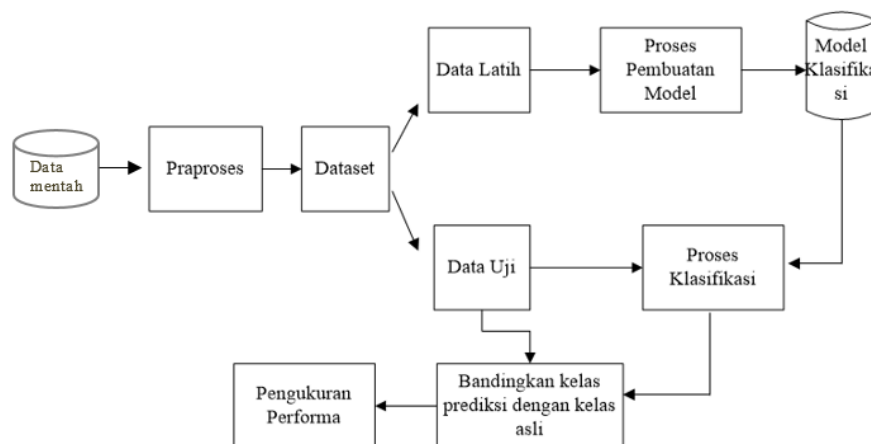
3. Recall

Recall merupakan kemampuan pengklasifikasi dalam memprediksi kelas instance yang benar-benar relevan diantara seluruh keseluruhan instance yang relevan.

$$\frac{TP}{(TP + FN)} \dots\dots\dots (3)$$

METODE

Penelitian ini dilakukan melalui beberapa tahap. Tahapan penelitian ini dapat dilihat pada diagram pada gambar 2 berikut ini.



Gambar 2. Metode Penelitian

Berikut merupakan penjelasan dari masing-masing tahap dalam penelitian ini.

1. Pengumpulan data

Data yang digunakan merupakan data mengenai produk kosmetik yang merupakan hasil produksi dari suatu perusahaan kosmetik. Data produk kosmetik yang digunakan adalah produk yang dikembalikan oleh konsumen dikarenakan berbagai alasan. Data produk yang digunakan pada penelitian ini terbatas pada produk yang diproduksi dan dikembalikan pada tahun yang sama sehingga usia produk kurang dari satu tahun. Total *instance* yang digunakan sebagai data penelitian ini sebanyak 432 *instance*.

2. Praproses data

Atribut dari data yang telah dikumpulkan dipilih sesuai dengan tujuan yang diharapkan. Atribut yang digunakan adalah nama produk dengan tipe atribut adalah text, kerusakan pada kemasan yang terdiri dari 11 atribut yang merupakan jenis kerusakan pada kemasan dengan tipe atribut adalah numerik, kerusakan pada isi yang terdiri dari 7 atribut yang merupakan jenis kerusakan pada isi produk dengan tipe atribut adalah numerik, dan atribut kelas yaitu kategori produk dengan tipe atribut adalah nominal. Atribut kategori produk merupakan atribut yang menjadi target klasifikasi. Kelas yang digunakan terdiri dari tiga kelas yaitu terbagi menjadi tiga kelas yaitu kuat, sedang, dan lemah. Pengkategorian produk berdasarkan 1) Produk yang masuk dalam kategori kuat adalah produk yang jumlah kerusakannya $\leq 1\%$ dari keseluruhan jumlah produk yang dikembalikan dalam satu tahun; 2) Produk yang masuk dalam kategori sedang adalah produk yang jumlah kerusakannya $> 1\%$ hingga $\leq 2\%$ dari keseluruhan jumlah produk yang dikembalikan dalam satu tahun; 3) Produk yang termasuk dalam kategori lemah merupakan produk yang jumlah kerusakannya $> 2\%$ dari keseluruhan jumlah produk yang dikembalikan dalam satu tahun. Dataset yang digunakan sebanyak 432 *instance* yang terbagi dalam tiga kelas yaitu kuat sebanyak 410 *instance*, sedang sebanyak 13 *instance*, dan lemah sebanyak 9 *instance*.

3. Klasifikasi

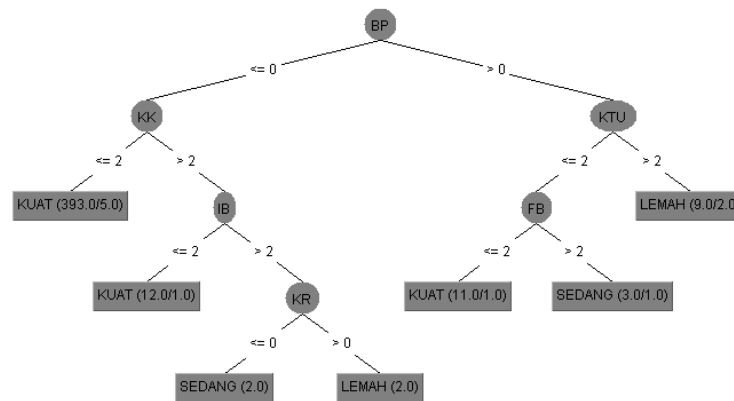
Pada tahap klasifikasi, dataset yang telah disiapkan sebelumnya dibagi menjadi dua dataset yaitu 1) Data latih dan 2) Data uji. Data latih digunakan untuk menghasilkan model klasifikasi, sedangkan data uji digunakan dalam evaluasi model klasifikasi. Pembagian jumlah dataset data latih dan data uji menggunakan metode *cross-validation* dengan 10 *folds*. Model klasifikasi dibuat dengan menggunakan algoritma pohon keputusan C4.5. Model klasifikasi ini digunakan dalam memprediksi kelas/kategori dari data uji.

4. Evaluasi model klasifikasi

Evaluasi model klasifikasi dilakukan untuk mengetahui performa dari model klasifikasi yang dihasilkan. Jika performa model klasifikasi telah sesuai dengan harapan, maka model klasifikasi dapat digunakan untuk memprediksi kelas dalam kasus nyata.

HASIL DAN PEMBAHASAN

Model klasifikasi diolah dengan menggunakan pohon keputusan C4.5 dan pembagian data latih dan data uji menggunakan metode *cross-validation* 10 *folds*. Model klasifikasi ini akan menjadi dasar dalam prediksi kasus nyata. Hasil model klasifikasi dapat dilihat pada gambar 3. Pada model tersebut terlihat bahwa atribut prediksi yang digunakan hanya enam atribut dari total atribut prediksi sebanyak 19 atribut. Atribut IB merupakan atribut untuk cacat isi, sedangkan sisanya (BP, KK, KTU, FB, dan KR) merupakan atribut dari jenis-jenis kerusakan pada kemasan produk. Pada model klasifikasi tersebut dapat terlihat bahwa kerusakan pada kemasan merupakan alasan utama mengapa suatu produk dikembalikan oleh konsumen. Hanya satu atribut kerusakan pada isi produk yang menjadi alasan suatu produk dikembalikan. Oleh karenanya, hal ini dapat menjadi evaluasi produk terutama dalam sisi kemasan yang kerusakannya bisa terjadi pada saat proses produksi, pengemasan, atau pada saat proses pengiriman produk.



Gambar 3. Model Klasifikasi dalam Bentuk Pohon Keputusan.

Hasil performa model klasifikasi yang dihasilkan (gambar 3) dalam bentuk *confusion matrix* sebagai berikut:

Tabel 2. *Confusion Matrix* Model Klasifikasi

		Kelas prediksi		
		Kuat	Sedang	Lemah
Kelas sebenarnya	Kuat	408	2	0
	Sedang	8	1	4
	Lemah	1	4	4

Pada *confusion matrix* pada gambar 3 terlihat bahwa kelas yang paling banyak diklasifikasikan benar adalah kelas kuat yang salah klasifikasinya hanya sebesar 0,49%. Kelas sedang mengalami paling banyak salah klasifikasi yaitu sebesar 92,31%. Hal ini menunjukkan bahwa model klasifikasi tidak mampu memprediksi kelas sedang, namun sangat baik dalam memprediksi kelas kuat. Hal ini terjadi karena jumlah *instance* masing-masing kelas sangat tidak seimbang (*imbalanced class distribution*). Model klasifikasi ini menghasilkan tingkat akurasi sebesar 95,60% dengan tingkat galat sebesar 4,4%. Hal ini dimungkinkan karena akurasi yang tinggi pada kelas kuat sehingga tingkat akurasi model klasifikasi secara keseluruhan juga tinggi walaupun banyak terjadi salah klasifikasi pada kelas lainnya. Model klasifikasi ini memiliki tingkat precision sebesar 0,943 dan recall sebesar 0,956. Tingginya precision dan recall walaupun salah klasifikasi pada kelas sedang dan lemah lebih dari 50% dikarenakan jumlah *instance* yang sangat tidak seimbang dengan kelas kuat memiliki proporsi sebesar 94,91% dari keseluruhan dataset.

KESIMPULAN

Data mengenai produk retur dapat diolah untuk menjadi informasi yang bermanfaat bagi pelaku bisnis. Salah satu metode pengolahan data yaitu klasifikasi dapat digunakan dalam pengolahan data produk retur untuk mengklasifikasikan produk retur ke dalam beberapa kelas. Dari hasil model klasifikasi yang dihasilkan oleh algoritma C4.5 dapat digunakan untuk pengembangan aturan pengembalian produk dan juga evaluasi dari proses produksi produk. Model klasifikasi yang dihasilkan pada penelitian ini dinilai memiliki performa yang cukup dengan tingkat akurasi sebesar 95,60% dengan tingkat galat sebesar 4,4%, precision sebesar 0,943 dan recall sebesar 0,956. Akan tetapi model klasifikasi ini perlu dievaluasi dan diperbaiki, karena rendahnya kemampuan pengklasifikasi dalam mengklasifikasi *instance* ke dalam kelas ‘sedang’ dan ‘lemah’. Hal ini dikarenakan proporsi jumlah *instance* dalam masing-masing kelas tidak memiliki jumlah yang seimbang sehingga menimbulkan bias terhadap salah satu kelas yang memiliki jumlah *instance* sangat banyak. Oleh karenanya diharapkan untuk memperbaiki proporsi jumlah *instance* dalam masing-masing kelas agar model klasifikasi dapat diterapkan dalam pengujian kasus nyata.

DAFTAR PUSTAKA

- [1] C. Bonifield, C. Cole dan R. L. Schultz, "Product returns on the internet: a case of mixed signals?," *Journal of Business Research*, 63(9-10), pp. 1058-1065, 2010.
- [2] E. J. Blocher, K. H. Chen, G. Cokins dan T. W. Lin, *Cost Management, Manajemen Biaya*, Jakarta: Salemba Empat, 2007.
- [3] Republik Indonesia, "Undang-Undang Perlindungan Konsumen No. 8 Tahun 1999," 1999. [Online].
Available: <http://sireka.pom.go.id/requirement/UU-8-1999-Perlindungan-Konsumen.pdf>.
- [4] P. Jennifer C. Kerr, "Retailers tracking what customers return," 12 Agustus 2013. [Online].
Available: <https://www.usatoday.com/story/money/business/2013/08/12/retailers-tracking-customers-returns/2642607/>.
- [5] AllBusiness.com, "The Importance of a Good Return Policy," 10 Juli 2007. [Online].
Available:
https://archive.nytimes.com/www.nytimes.com/allbusiness/AB4353479_primary.html?mcubz=1&pagewanted=all.
- [6] S.-P. Jeng, "Increasing customer purchase intention through product return policies: The pivotal impacts of retailer brand familiarity and product categories," *Journal of Retailing and Consumer Services*, 39, pp. 182-189, 2017.
- [7] N. Lakshmi, T. S. Indumathi dan N. Ravi, "A Study on C. 5 decision tree classification algorithm for risk predictions during pregnancy," *Procedia Technology*, 24, pp. 1542-1549, 2016.
- [8] Hssina, A. Merbouha, H. Ezzikouri dan M. Erritali, "A comparative study of decision tree ID3 and C4.5," *International Journal of Advanced Computer Science and Applications*, 4(2), 2014.
- [9] Mulyadi, *Sistem Akuntansi, Edisi Tiga*, Jakarta: Salemba Empat, 2001.
- [10] Soemarso, *Akuntansi Suatu Pengantar. Buku ke 2. Edisi 5*, Jakarta: Salemba Empat, 2009.
- [11] P. N. Tan, M. Steinbach dan V. Kumar, *Introduction to Data Mining*, Boston: Pearson Addison Wesley, 2005.
- [12] R. Nisbet, J. Elder dan G. Miner, *Handbook of Statistical Analysis and Data Mining Applications*, Elsevier, Inc, 2009.
- [13] J. Han, M. Kamber dan J. Pei, *Data mining: concepts and techniques*, third edition, Morgan Kauffmann, 2012.
- [14] S. B. Kotsiantis, I. Zaharakis dan P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerging artificial intelligence applications in computer engineering*, 160, pp. 3-24, 2007.
- [15] J. R. Quinlan, *C4.5: programs for machine learning*, Elsevier, 2014.
- [16] T.-S. Lim, W.-Y. Loh dan Y.-S. Shih, "A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms," *Machine learning*, 40(3), pp. 203-228, 2000.
- [17] V. Kotu dan B. Deshpande, *Predictive analytics and data mining: concepts and practice with rapidminer*, Massachusetts: Morgan Kaufmann, 2015.