

Penerapan Algoritma C4.5 dan Random Forest pada Tingkat Penjualan Serum Somethinc di Shopee*

Rismayanti^{1†}, Muhammad Nur Aidi¹, Hari Wijayanto¹

¹Department of Statistics, IPB University, Indonesia

*corresponding author: rismayanti_r@apps.ipb.ac.id

Copyright © 2023 Rismayanti, Muhammad Nur Aidi, Hari Wijayanto. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Online buying and selling activities in Indonesia are increasing. Shopee is an online buying and selling platform with the highest visits in Indonesia in the fourth quarter of 2022. The category with the highest transactions at Shopee is beauty products. Somethinc is a very successful local beauty product at Shopee which have highest sales of serum products in Indonesia. This study applies the classification method C4.5 and Random Forest to see important variables in the sales of Somethinc serum at Shopee. The variables used come from store profiles which include: number of followers, number of products, chat performance, store rating, and length of stay. Continuous sales data is discretized using k-means into ordinal data with low, medium, and high levels. There is an imbalance of data in the sales class so that the SMOTE technique is used. The C4.5 algorithm produces a decision tree that contains rules for classification. Random Forest generates the order of variable importance based on the Mean Decrease Gini (MDG) values in descending order, which are as follows: number of followers, number of products, message performance, joining time, and store rating.

Keywords: C4.5, discretization, online shop, Random Forest, SMOTE.

1. Pendahuluan

Kegiatan belanja online sudah banyak dilakukan oleh masyarakat Indonesia. Pergeseran tren belanja dari offline ke online membuat bisnis online menjadi pilihan baru konsumen. Belanja online dinilai lebih praktis dan fleksibel sehingga memudahkan pelanggan untuk melakukan transaksi. Shopee merupakan salah satu e-commerce yang populer digunakan di Indonesia. Berdasarkan databoks (2022), jumlah pengunjung bulanan Shopee pada kuartal IV tahun 2022 menduduki posisi pertama dengan 191,6 juta pengunjung.

Pertumbuhan belanja online didominasi oleh produk kecantikan. Data dari Compas (2022) menunjukkan bahwa produk kecantikan menduduki posisi pertama penjualan tertinggi di e-commerce pada paruh pertama tahun 2022. Serum wajah

* Received: Jun 2023; Reviewed: Sep 2023; Published: Dec 2022

menjadi produk kecantikan yang memiliki tingkat penjualan tertinggi ke dua di shopee. Pertumbuhan tingkat belanja online pada produk kecantikan, khususnya produk serum wajah diiringi juga dengan pertumbuhan industri kecantikan dalam negeri. Brand lokal telah mampu bersaing dan mendominasi pasar produk serum wajah di Indonesia. Compas (2022) telah memeringkatkan 10 brand serum wajah dengan tingkat penjualan tertinggi, 8 diantaranya merupakan brand lokal. Somethinc merupakan salah satu brand kecantikan lokal yang berdiri sejak tahun 2019 dengan produk pertamanya yaitu serum wajah yang kini telah sukses di pasaran dan menduduki posisi pertama penjualan tertinggi di e-commerce.

Menurut Wibowo (2015), faktor penting dalam kesuksesan bisnis online adalah membangun kepercayaan dan loyalitas pelanggan. Pada bisnis online, pembeli tidak dihadapkan langsung dengan barang yang akan dibeli, sehingga membangun kepercayaan dalam bisnis online berbeda dengan bisnis offline. Informasi profil toko online di Shopee memudahkan pelanggan untuk menilai kemudian mengambil keputusan pembelian karena informasi profil toko ini menunjukkan reputasi yang telah dibangun suatu toko online. Informasi profil toko online yang ditampilkan oleh Shopee meliputi jumlah produk, jumlah pengikut, performa chat, penilaian/rating, dan usia toko. Nadine et al. (2020) melakukan penelitian mengenai analisis pengaruh profil toko terhadap tingkat penjualan, analisis dilakukan secara eksploratif. Hasil dari penelitian tersebut adalah terdapat pengaruh jumlah followers dan rating toko terhadap penjualan.

Penelitian ini menggunakan profil toko sebagai peubah prediktor dan penjualan sebagai peubah respon. Data penjualan berupa data numerik. Diskritisasi data dilakukan pada penelitian ini untuk dapat mempermudah interpretasi dan analisis pada setiap tingkat penjualan Serum Somethinc. Diskritisasi data menghasilkan peubah diskrit yang akan digunakan untuk analisis klasifikasi. Analisis klasifikasi dilakukan pada penelitian ini menggunakan Algoritma C4.5 dan Random Forest. Algoritma C4.5 menghasilkan pohon keputusan yang membentuk aturan dalam proses klasifikasi, sementara Random Forest dapat memeringkatkan kepentingan peubah dengan menggunakan ukuran MDG (Mean Decrease Gini). Data yang tak seimbang dapat memengaruhi hasil klasifikasi sehingga hasilnya cenderung memihak kepada kelas mayoritas (Shyntia et al. 2022). Synthetic Minority Over Sampling Technique (SMOTE) dapat mengatasi masalah data tak seimbang dengan membangkitkan data sintetis pada kelas minoritas.

Penelitian sebelumnya mengenai algoritma C4.5 dilakukan oleh Robby et. al. (2019) untuk menentukan faktor yang memengaruhi munculnya profesional blogger. Penelitian mengenai Random Forest dilakukan oleh Mayapada et al. (2022) untuk mengetahui faktor yang memengaruhi indeks pembangunan manusia di Indonesia. Pamuji dan Ramadhan (2021) melakukan komparasi antara metode Random Forest dan Decision Tree untuk memprediksi keberhasilan Immunotherapy. Hasil penelitian tersebut menunjukkan metode Random Forest memiliki nilai akurasi yang lebih baik dibandingkan metode Decision Tree.

2. Metodologi

2.1 Data

Data yang digunakan merupakan data penjualan selama bulan Oktober 2022. Data reseller resmi berasal dari halaman web resmi Somethinc di <https://somethinc.com/id/stockist>, sementara data profil setiap toko *online* berasal dari halaman profil toko *online* di Shopee yang diambil pada 31 Oktober 2022. Data terdiri

dari 5 peubah prediktor dengan jumlah amatan sebanyak 338 toko *online* yang merupakan reseller resmi dari Somethinc. Peubah yang digunakan terdapat pada Tabel 1.

Tabel 1. Peubah yang digunakan dalam penelitian

Kode	Peubah	Keterangan
Y	sales	Jumlah penjualan pada bulan Oktober 2022
X1	n.prod	Banyaknya jenis produk yang tersedia di toko
X2	n.followers	Banyaknya pengikut toko
X3	chat.perform	Performa pesan berdasarkan persentase pesan yang dibalas oleh toko
X4	rating	Rataan penilaian pelanggan terhadap toko
X5	l.stay	Usia toko di Shopee

2.2 Diskritisasi

Diskritisasi merupakan transformasi peubah kontinyu menjadi peubah diskrit. Analisis cluster adalah salah satu metode yang paling populer untuk diskritisasi. Penelitian ini menggunakan metode k-means untuk diskritisasi data. Metode k-means adalah metode berbasis centroid. Proses pengelompokan data menjadi sebuah cluster dapat dilakukan dengan menghitung jarak terdekat dari data ke centroid. Jarak yang digunakan dalam penelitian ini adalah jarak Euclidian dengan rumus:

$$dist(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

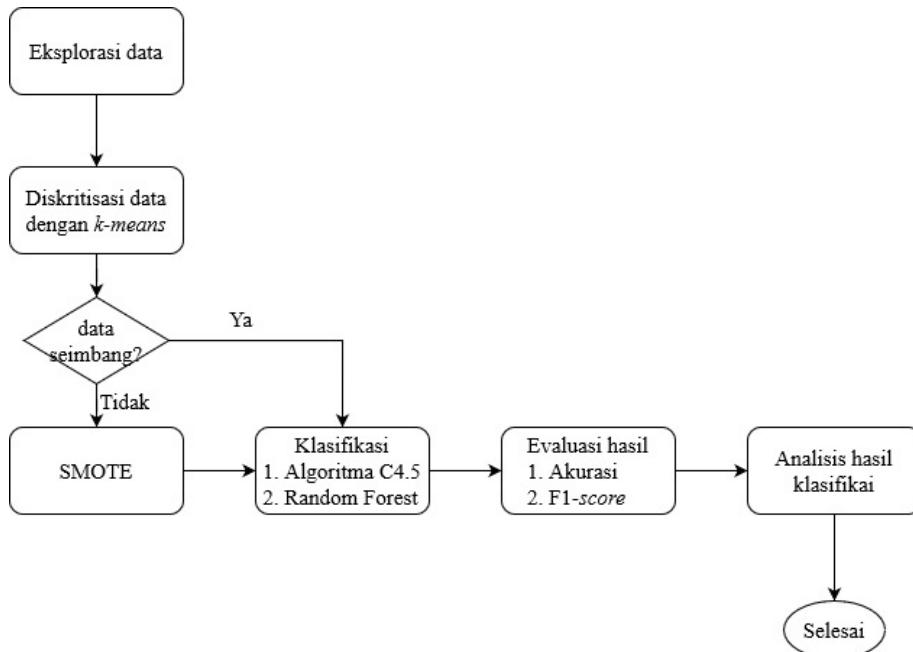
2.3 Algoritma C4.5

Algoritma C4.5 merupakan modifikasi dari algoritma ID3. Metode C4.5 merupakan teknik pembentukan pohon keputusan yang menghitung nilai gain. Nilai gain tertinggi sebagai simpul pertama atau simpul akar (Sudrajat et. al. 2017). Output dari algoritma C4.5 adalah sebuah pohon keputusan atau sering dikenal dengan decision tree. Kelebihan dari algoritma C4.5 yaitu mudah dimengerti, fleksibel dan menarik karena mampu divisualisasikan dalam gambar (Ente et. al. 2020).

2.4 Random Forest

Metode klasifikasi Random Forest terdiri dari sekelompok pohon klasifikasi tunggal yang digabungkan menjadi satu model. Random Forest menghasilkan kombinasi pohon yang baik dari setiap model pohon keputusan, lalu digabungkan menjadi satu model. Setiap pohon dibuat dengan menggunakan sebagian peubah prediktor secara acak dari seluruh peubah yang ada, sehingga menghasilkan pohon klasifikasi yang independen dan memiliki korelasi yang rendah. Penentuan klasifikasi dengan Random Forest diambil berdasarkan hasil voting dari tree yang terbentuk (Breiman 2001).

Terdapat dua ukuran kepentingan yang bisa digunakan dalam menentukan peubah penting pada Random Forest. Mean Decrease Accuracy (MDA) dan Mean Decrease Gini (MDG) merupakan ukuran kepentingan yang dapat memeringkatkan setiap peubah yang digunakan dalam Random Forest. MDG memiliki hasil yang lebih robust dibandingkan dengan MDA dalam memeringkatkan peubah penting (Calle dan Urrea 2010). MDG merupakan rasio antara penjumlahan dari impuritas tereduksi yang disebabkan oleh peubah prediktor ke-s dengan banyak pohon yang terbentuk. Semakin tinggi nilai MDG, semakin penting peran peubah prediktor dalam



pembentukan pohon klasifikasi (Dito et al. 2019).

2.5 Metode Penelitian

Berdasarkan Gambar 1, tahapan penelitian adalah sebagai berikut

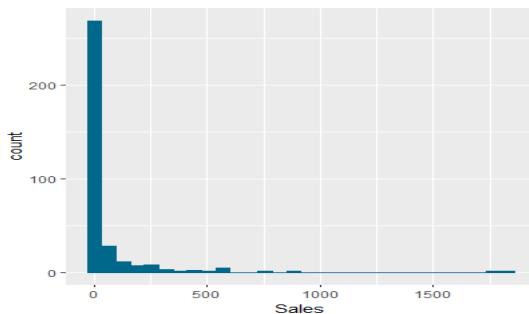
1. Melakukan eksplorasi data untuk melihat gambaran umum data.
2. Melakukan diskritisasi pada data. Penelitian ini menggunakan metode *k-means* untuk diskritisasi data. Metode *k-means* adalah metode berbasis *centroid*. Proses pengelompokan data menjadi sebuah cluster dapat dilakukan dengan menghitung jarak terdekat dari data ke *centroid*. Jarak yang digunakan dalam penelitian ini adalah jarak *Euclidian*.
3. Mengatasi ketidakseimbangan data dengan SMOTE.
4. Menerapkan metode klasifikasi
 - a. Algoritma C4.5 merupakan modifikasi dari algoritma ID3. Metode C4.5 merupakan teknik pembentukan pohon keputusan dengan menghitung nilai Gain.
 - b. *Random Forest* terdiri dari sekelompok pohon klasifikasi tunggal yang digabungkan menjadi satu model. Pembentukan pohon klasifikasi pada *Random Forest* dilakukan dengan menghitung nilai Impuritas.
5. Mengevaluasi hasil klasifikasi dengan melihat nilai akurasi dan F1-score.

Gambar 1 Diagram alur metode penelitian

3. Hasil dan Pembahasan

3.1 Karakteristik Data

Penjualan serum somethinc pada setiap toko *online* selama bulan Oktober menyebar dari 0 hingga 1829. Dari 338 toko *online*, sebanyak 148 toko *online* tidak memiliki penjualan selama bulan Oktober. Gambar 2 merupakan histogram penjualan. Data penjualan menumpuk pada sisi kiri dan memiliki kesenjangan tingkat penjualan antar toko *online* sehingga akan sulit dianalisis. Maka dari itu, data penjualan yang kontinyu ini perlu dilakukan diskritisasi untuk mempermudah proses analisis data.



Gambar 2. Histogram data penjualan

Tabel 2 menunjukkan statistic deskriptif pada setiap peubah yang digunakan. Peubah Y, X1, dan X2 memiliki nilai standar deviasi yang sangat tinggi. Peubah performa pesan (X3) memiliki nilai minimal yang sangat kecil, namun memiliki nilai rataan dan median yang cukup besar yaitu 0,75 dan 0,81. Artinya banyak toko *online* yang sudah memberikan layanan yang baik dalam membela pesan. Peubah rating (X4) berada pada rentang 4,5-5 dengan rataan dan median yang sama yaitu 4,9. Hal ini menunjukkan bahwa persaingan di e-commerce cukup ketat. Sementara peubah waktu bergabung (X5) menyebar dari 0-8 tahun.

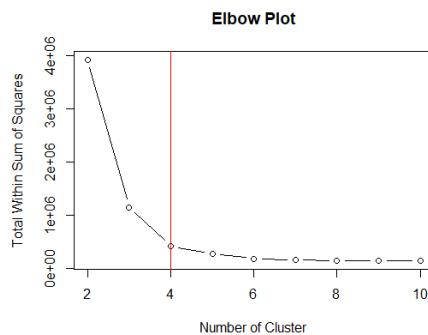
Tabel 2. Statistik deskriptif

Peubah	Rataan	Median	Min	Maks	Std
Y	53,01	1,00	0,00	1829	176,72
X1	651,46	273,50	1,00	6400	940,22
X2	50858,70	4200,00	4,00	3300000	212029,40
X3	0,75	0,81	0,04	1	0,22
X4	4,90	4,90	4,50	5	0,06
X5	3,88	4,00	0,00	8	2,04

3.2 Diskritisasi data

Diskritisasi data pada penelitian ini menggunakan metode *k-means*. Langkah pertama dalam metode *k-means* adalah menentukan jumlah kelompok yang akan digunakan. Penentuan jumlah kelompok *k-means* ini dilakukan dengan menghitung nilai *Within Cluster Sum of Squared* (WCSS) pada setiap k kelompok mulai dari 2-10. Gambar 3 menunjukkan nilai SSE pada 2 kelompok dan 3 kelompok berbeda cukup signifikan. Begitupun nilai SSE pada 3 kelompok dan 4 kelompok. Namun, pada 4 kelompok dan seterusnya menunjukkan perubahan nilai WCSS yang sudah tidak signifikan lagi. Maka

pada penelitian ini, jumlah kelompok yang digunakan untuk *k-means* adalah 4 kelompok.



Gambar 3. WCSS pada setiap jumlah kelompok *k-means*

Tabel 3 menunjukkan hasil diskritisasi pada peubah Y dengan *k-means*. Terdapat 4 kelompok yang masing-masing diberikan label rendah, sedang, tinggi, dan sangat tinggi. Hanya 2 toko yang masuk kedalam kelompok penjualan sangat tinggi, maka kemudian kelompok amatan ini dilakukan analisis secara terpisah dan tidak diikutsertakan ke dalam model klasifikasi.

Tabel 3. Hasil diskritisasi dengan *k-means*

Label	Rentang nilai	n
Rendah	0-100	297
Sedang	101-400	28
Tinggi	401-1000	11
Sangat tinggi	>1000	2

3.3 Klasifikasi

Setelah diskritisasi, penelitian dilanjutkan dengan proses klasifikasi. Penelitian ini juga menerapkan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) dikarenakan terdapat ketidakseimbangan pada data.

3.3.1 Algoritma C4.5

Performa hasil klasifikasi menggunakan algoritma C4.5 dapat dilihat pada Tabel 4. Akurasi hasil algoritma C4.5 sebelum SMOTE adalah 87,78% dengan F1-score untuk masing-masing kelas yaitu 94,86% untuk kelas 'rendah', 64,71% untuk kelas 'sedang', dan 40% untuk kelas 'tinggi'. Berdasarkan nilai akurasinya, performa algoritma C4.5 sudah cukup baik, namun dilihat dari performa F1-score masih terjadi ketimpangan pada setiap kelas data.

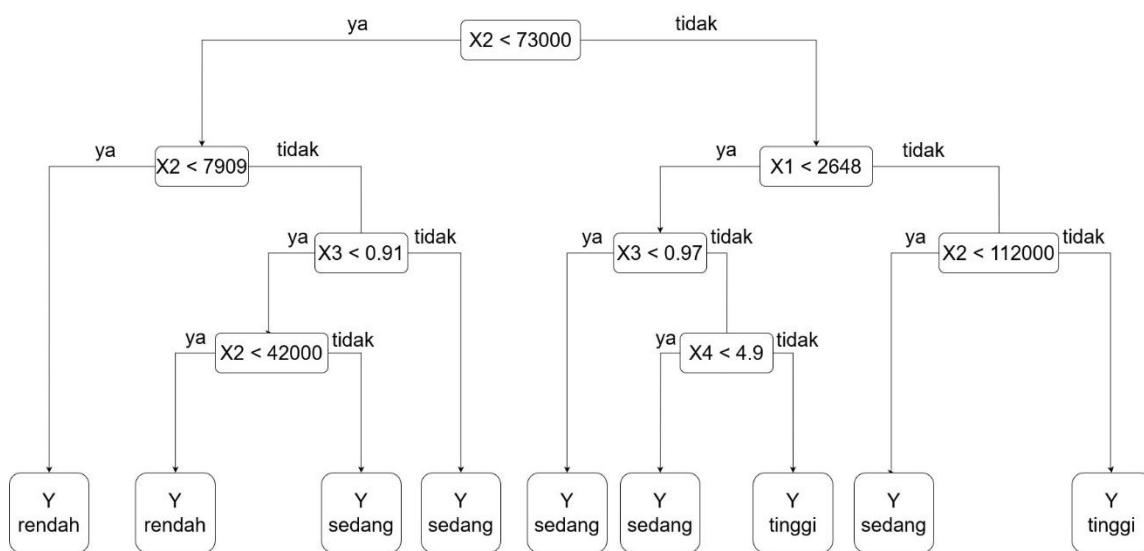
Performa algoritma C4.5 untuk data yang telah dilakukan SMOTE mengalami peningkatan pada nilai akurasi dan keseimbangan pada F1-score. Akurasi algoritma C4.5 pada data SMOTE yaitu 89,49%, dengan hasil F1-score untuk setiap kelas 91,62% untuk kelas 'rendah', 83,66% untuk kelas 'sedang', dan 93,21% untuk kelas 'tinggi'.

Tabel 4 Performa model algoritma C4.5 tanpa SMOTE dan dengan SMOTE

	Tanpa SMOTE	SMOTE
Akurasi	0,8778	0,8949
F1-score 'rendah'	0,9486	0,9162
F1-score 'sedang'	0,6471	0,8366
F1-score 'tinggi'	0,4000	0,9321

Pohon keputusan yang dihasilkan dari algoritma C4.5 dapat dilihat pada Gambar 4. Aturan yang terbentuk dari pohon keputusan adalah sebagai berikut :

1. Penjualan rendah banyak terjadi pada toko dengan jumlah pengikut kurang dari 7909 atau toko dengan jumlah pengikut berada pada rentang 7909-42000, namun performa pesan kurang dari 0,91.
2. Penjualan sedang banyak terjadi pada:
 - a. Jumlah pengikut lebih 7909 dan performa pesan lebih dari 0,91.
 - b. Performa pesan kurang dari 0,91 namun jumlah pengikut lebih dari 42000, dan kurang dari 73000.
 - c. Jumlah pengikut lebih dari 73000, jumlah produk kurang dari 2648, dan performa pesan kurang dari 0,97.
 - d. Jumlah pengikut lebih dari 73000, jumlah produk kurang dari 2648, performa pesan lebih dari 0,97 dan rating kurang dari 4,9.
 - e. Jumlah pengikut berada pada rentang 73000-112000 dan jumlah produk 2648.
3. Penjualan tinggi banyak terjadi pada toko dengan jumlah pengikut lebih dari 73000, jumlah produk lebih dari 2648, performa pesan lebih dari 0,97, dan rating lebih dari 4,9 atau toko dengan jumlah pengikut lebih dari 112000 dan jumlah produk lebih dari 2648.



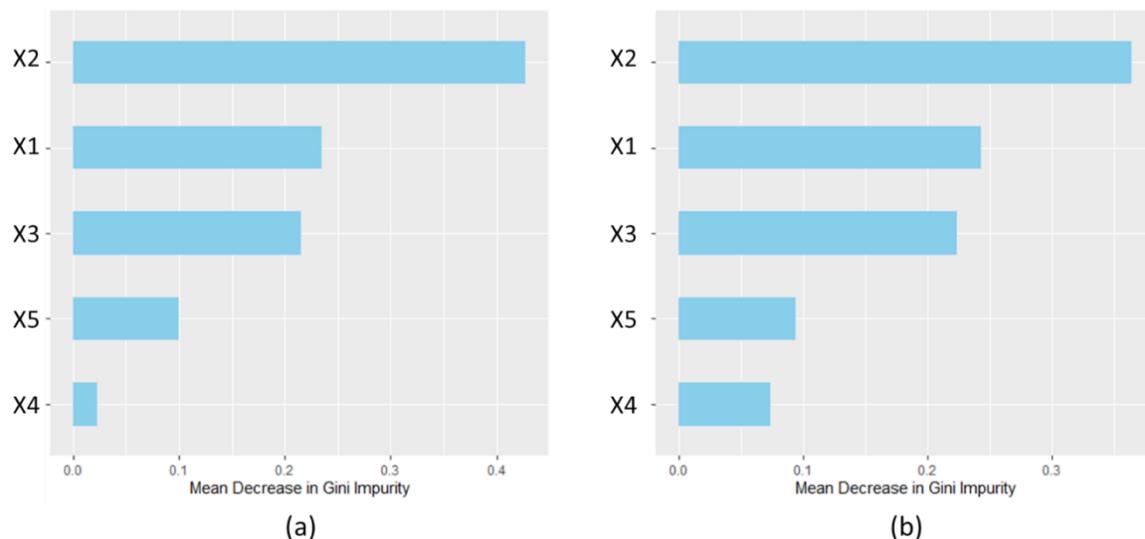
Gambar 4 Pohon keputusan algoritma C4.5

Secara umum, hasil pohon keputusan algoritma C4.5 menunjukkan bahwa penjualan tinggi banyak terjadi pada kondisi jumlah pengikut yang banyak, jumlah produk yang banyak, performa pesan yang sangat baik, dan rating yang tinggi,

Sementara penjualan rendah banyak terjadi pada jumlah pengikut yang rendah dan performa pesan yang rendah.

3.3.2 Random Forest

Hasil klasifikasi dengan menggunakan *Random Forest* dapat dilihat pada Tabel 5. Sebelum penerapan SMOTE, model *Random Forest* menghasilkan akurasi sebesar 89,95% dan F1-score untuk kelas ‘rendah’, ‘sedang’, dan ‘tinggi’ berturut-turut 96,07%, 20,51%, dan 43,48%. Nilai akurasi hasil *Random Forest* sudah cukup baik, namun nilai F1-score pada kelas minoritas masih sangat rendah, sehingga hasil klasifikasi *Random Forest* masih cenderung memihak pada kelas mayoritasnya.



Gambar 5 Kepentingan peubah berdasarkan ukuran Mean Decrease Gini (MDG): (a) Data tanpa SMOTE. (b) Data dengan SMOTE

Performa Pemodelan *Random Forest* mengalami peningkatan menjadi 96,39% setelah dilakukan SMOTE pada data. Hasil F1-score juga mengalami peningkatan secara keseluruhan terutama pada kelas minoritasnya. F1-score pada kelas ‘rendah’, ‘sedang’, dan ‘tinggi’ secara berurutan adalah 96,56%, 95,24%, dan 98,23%. Peningkatan ini menunjukkan bahwa penerapan teknik SMOTE dapat mengatasi ketidakseimbangan data dan mampu meningkatkan performa klasifikasi *Random Forest*.

Tabel 5 Perbandingan performa *Random Forest* sebelum SMOTE dengan sesudah SMOTE

	Tanpa SMOTE	SMOTE
Akurasi	0,8995	0,9639
F1-score 'rendah'	0,9607	0,9656
F1-score 'sedang'	0,2051	0,9524
F1-score 'tinggi'	0,4348	0,9823

Algoritma *Random Forest* menghasilkan tingkat kepentingan dari peubah prediktor. Tingkat kepentingan peubah ini berdasarkan nilai *Mean Decrease Gini* (MDG). Gambar 5 menunjukkan hasil *Mean Decrease Gini* pada setiap peubah

prediktor. Secara umum, hasil MDG pada *Random Forest* sebelum dan sesudah SMOTE menghasilkan urutan kepentingan peubah yang sama. Terdapat 3 peubah yang memiliki nilai kepentingan yang tinggi, diatas 0,2. Peubah tersebut yaitu jumlah pengikut (X2), jumlah produk (X1), dan performa pesan (X3). Pentingnya peubah X2, X1, dan X3 menunjukkan bahwa peubah tersebut memiliki hubungan yang kuat dengan peubah respon dalam model klasifikasi ini. Peubah-peubah ini memberikan informasi yang relevan dan mempengaruhi pengambilan keputusan oleh model, sementara peubah usia toko (X5) dan rating (X4) memiliki nilai kepentingan yang rendah, bernilai dari 0,1 hingga 0,02.

3.4 Evaluasi

Performa klasifikasi menggunakan algoritma C4.5 dan Random Forest dapat dilihat pada Tabel 6. Meskipun Random Forest dan Algoritma C4.5 memiliki performa klasifikasi yang berbeda, namun keduanya memiliki performa yang cukup baik.

Tabel 6 Performa metode klasifikasi Algoritma C4.5 dan Random Forest

	RF	C4.5
Akurasi	0.9639	0.8949
F1-score 'rendah'	0.9656	0.9162
F1-score 'sedang'	0.9524	0.8366
F1-score 'tinggi'	0.9823	0.9321

Random Forest menghasilkan peubah penting yang terdiri dari jumlah pengikut, jumlah produk, dan performa pesan. Sementara, algoritma C4.5 menghasilkan pohon keputusan yang membentuk aturan klasifikasi dan menunjukkan karakteristik setiap kelas. Kelompok toko dengan tingkat penjualan rendah banyak terjadi pada toko yang memiliki jumlah pengikut sedikit dan performa pesan rendah. Kelompok toko dengan penjualan tinggi banyak terjadi pada toko yang memiliki jumlah pengikut yang banyak, jumlah produk yang banyak, performa pesan yang baik, dan rating yang tinggi. Sebagai upaya meningkatkan tingkat penjualan di toko *online*, hal yang perlu diperhatikan adalah :

1. Meningkatkan jumlah pengikut

Menurut Novitasari dan Redyanita (2021), strategi yang dapat dilakukan untuk meningkatkan trafik di Shopee adalah optimasi toko, iklan, optimasi produk, dan promosi. Optimasi toko merupakan upaya untuk membuat tampilan toko *online* menjadi lebih menarik dan mempermudah pelanggan dalam mencari produk. Iklan dan promosi dapat dilakukan untuk meningkatkan *brand awareness* sehingga akan meningkatkan trafik dan pengikut toko *online*.

2. Meningkatkan performa pesan

Pesan merupakan salah satu media untuk menghubungkan penjual dan pembeli. Penjual yang responsif dan dapat berkomunikasi dengan baik dapat meningkatkan kepuasan pelanggan sehingga akan meningkatkan peluang agar pelanggan dapat terus berbelanja atau *repurchase* produk.

3. Menambah variasi/kelengkapan produk

Toko dengan produk yang lebih lengkap biasanya lebih digemari karena pelanggan dapat berbelanja sekaligus dalam satu toko dan dapat menghemat ongkos kirim.

4. Simpulan

Data penjualan serum Somethinc dikelompokan menggunakan k-means menjadi 4 tingkatan penjualan, yaitu: penjualan rendah, sedang, tinggi, dan sangat tinggi. Terjadi ketidakseimbangan data pada kelas penjualan sehingga diterapkan teknik SMOTE. Penerapan SMOTE ini meningkatkan performa pada kedua metode klasifikasi yang digunakan (C4.5 dan Random Forest).

Algoritma C4.5 menghasilkan pohon keputusan yang membentuk aturan klasifikasi. Kelompok toko dengan penjualan tinggi banyak terjadi pada toko yang memiliki jumlah pengikut yang banyak, jumlah produk yang banyak, performa pesan yang baik, dan rating yang tinggi. Random Forest menghasilkan peringkat kepentingan peubah berdasarkan nilai MDG, peubah yang memiliki tingkat kepentingan dari tinggi ke rendah berturut-turut : jumlah pengikut, jumlah produk, performa pesan, waktu bergabung, dan rating. Faktor penting dalam tingkat penjualan Serum Somethinc di Shopee berdasarkan Algoritma C4.5 dan Random Forest yaitu jumlah pengikut, jumlah produk, dan performa pesan.

Daftar Pustaka

- Compas. 2022. Compas Market Insight: Report 1 Semester 2022 [internet]. [diacu pada September]. Tersedia dari: <https://compas.co.id/market-insight-report/>.
- Databoks. 2022. 10 E-Commerce dengan Pengunjung Terbanyak Kuartal I 2022 [internet]. [diacu pada September 2022]. Tersedia dari: <https://databoks.katadata.co.id/datapublish/2022/07/19/10-e-commerce-dengan-pengunjung-terbanyak-kuartal-i-2022/>.
- Dito GA, Safitri A, Afendi FM, Anisa R, Salim A, Sartono B. 2019. Graphical User Interface (GUI) for the Least Absolute Shrinkage and Selection Operator (LASSO) Regression. IOP Conference Series: Earth and Environmental Science. 299(1): 012031. doi:10.1088/1755-1315/299/1/012031.
- Ente, DR, Thamrin SA, Kuswanto H, Arifin S, Andreza. 2020. Klasifikasi Faktor- Faktor Penyebab Penyakit Diabetes Melitus di Rumah Sakit Unhas Menggunakan Algoritma C4.5. Indonesian Journal of Statistics and Its Applications. 4(1): 80-88.
- Mayapada R, Yanti RW, Syarifuddin S. 2022. Analisis Tingkat Kepentingan terhadap Faktor-Faktor yang Memengaruhi Indeks Pembangunan Manusia di Indonesia. Journal of Mathematics Theory and Applications. 4(2): 45-49.
- Nadine JA, Jauhar I, Rahmarani AR, Rakhmawati NA. 2020. Analisis Pengaruh Profil Toko Resmi Smartphone pada Situs E-Commerce terhadap Tingkat Penjualan. Jurnal Sistem dan Informatika. 15(1): 81-87.
- Pamuji FY, Ramadhan VP. 2021. Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotherapy. Jurnal Teknologi dan Manajemen Informatika. 7(1): 46-50.
- Robby, Lavinia, Nababan D. 2019. Analisa Algoritma C4.5 dalam Menentukan Faktor yang Memengaruhi Munculnya Professional Blogger. Jurnal Ilmiah Informatika. 7(1): 11-17.

Shyntia, F. D., Kurnia, A. and Dito, G. A. (2022) Pendekatan Metode CHAID dan Regresi Logistik dalam Menganalisis Faktor Berpengaruh pada Kejadian Stunting di Provinsi Jawa Barat, Xplore: Journal of Statistics, 11(1), pp. 48–58. doi: 10.29244/xplore.v11i1.857.

Sudrajat R, Irianingsih I, and Krisnawan D, “Analysis of data mining classification by comparison of C4. 5 and ID algorithms,” in IOP Conference Series: Materials Science and Engineering, vol. 166. IOP Publishing, 2017, pp.1–8.

Wibowo, Radetya Agung. 2015. Kesuksesan E-Commerce (Online Shopping) Melalui Trust dan Customer Loyalty. Jurnal Ekonomi Bisnis. 20(1): 1-52.on method for poverty alleviation at a sub-district level. *International Journal of Applied Mathematics and Statistics™*, 53(6): 77–84.