

Pemanfaatan Chi Square dan Ensemble Tree Classifier pada Model SVM, KNN dan C4.5 dalam Penjualan Online

Prastika Indriyanti¹, Wawan Gunawan²
^{1,2}Department of Informatic, Universitas Mercu Buana

Article Info

Article history:

Received 18 Jun 2024

Revised 19 Sept 2024

Accepted 1 Oct 2024

Keywords:

Feature selection
classification
ensemble
tree classifier
UMKM

ABSTRACT

This research aims to assist MSMEs in overcoming problems in online sales. Currently, sellers only prepare stock, unaware of the product's sales performance in their market segment. In the city of Tangerang alone, there are 222,602 MSMEs with various product categories. Therefore, besides utilizing offline sales, business actors should also engage in online sales. Currently, sellers only prepare stock, unaware of the product's sales performance in their market segment. The SVM, KNN, and C4.5 algorithms are used to build prediction models based on the selected features. Using feature selection, it was found that the influential features are Estimated Shipping Cost, Shipping Cost Paid by Buyer, Total Product Price, and Estimated Shipping Cost Discount. Three algorithms—SVM, KNN, and C4.5—were used in the tests. The C4.5 model with data from the ensemble tree classifier and six features had the best accuracy rate (0.86%). The C4.5 model with ten features, KNN with six features, and KNN with ten features all utilize data from the ensemble tree classifier and achieve an accuracy rate of 0.85%.

Copyright © 2024 Universitas Indraprasta PGRI.
All rights reserved.

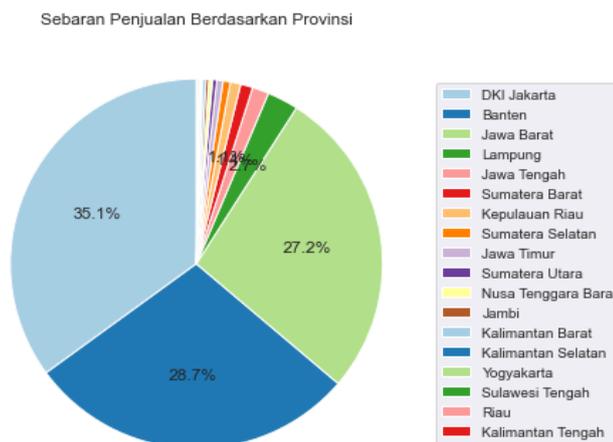
Corresponding Author:

Prastika Indriyanti,
Department of Informatic,
Universitas Mercu Buana,
Jl. Meruya Selatan No.1 Kembangan, Jakarta Barat.
Email: prastika.indriyanti@mercubuana.ac.id

1. PENDAHULUAN

Penjualan makanan secara online telah menjadi fenomena yang tidak bisa diabaikan dalam industri kuliner modern. Banyak sekali pelaku usaha yang menjual makanan baik secara offline maupun dilakukan secara online. Saat ini saja pelaku UMKM yang tersebar di Indonesia sebanyak 66 juta atau meningkat 1,52% dari tahun sebelumnya [1] dan tertinggi berada di provinsi Jawa Barat sekitar 1,4 juta, sehingga pelaku usaha harus memiliki strategi sendiri untuk dapat menembus pasar di luar daerahnya jika ingin tetap bertahan, khususnya target generasi Z [2]. Untuk daerah Kota Tangerang saja terdapat 222.602 UMKM dengan berbagai macam kategori produk yang dijual [3], oleh karena itu selain memanfaatkan penjualan secara offline, pelaku usaha juga hendaknya melakukan penjualan secara online. Saat ini di Indonesia sebanyak 3,79 juta UKM sudah melakukan transformasi digital menuju online [4], dan di Kota Tangerang sudah mencapai 33.339 UMKM di tahun 2022 [5].

Permasalahan yang terjadi saat ini para penjual hanya sebatas melakukan penyiapan stok barang saja tanpa mengetahui sejauh mana produk tersebut terjual untuk segmen pasar yang mana, sehingga sangat sulit untuk meningkatkan omset penjualan. Selain itu sulitnya bersaing dengan para generasi Z yang memang sangat faham berbagai trik dan teknik dalam memahami aspek pasar yang ada dengan kecanggihan teknologi yang dimiliki oleh mereka [6], [7]. Selain itu dengan pemanfaatan teknologi seharusnya target pasar dapat menjadi lebih luas, namun pada kenyataannya saat ini target pembeli hanya di provinsi yang berdekatan dengan lokasi toko berada dengan sebaran pelanggan dapat dilihat pada gambar 1.



Gambar 1 Sebaran penjualan berdasarkan Provinsi

Berdasarkan permasalahan tersebut, salah satu cara untuk mengatasi tantangan ini adalah dengan menerapkan algoritma klasifikasi yang membantu mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan pola dalam data [8], [9]. Algoritma klasifikasi membantu mengidentifikasi jenis makanan dan mengelompokkannya ke dalam kategori yang sesuai. Bagaimana algoritma klasifikasi dapat membantu untuk melakukan segmentasi pelanggan [10]–[12] yang berbeda berdasarkan perilaku belanja, preferensi produk, dan karakteristik demografis. Bagaimana memprediksi perilaku pembelian masa depan pelanggan berdasarkan pola pembelian sebelumnya [13]–[15], sehingga membantu bisnis dalam merencanakan strategi penjualan dan pemasaran.

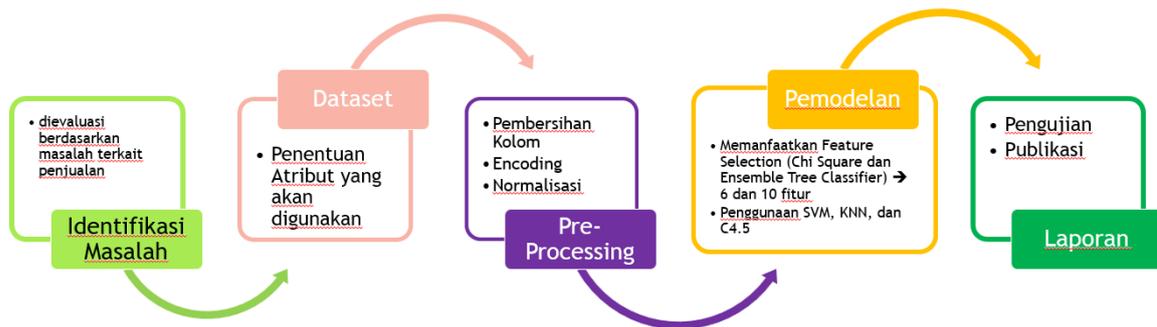
Penelitian sebelumnya membandingkan algoritma K-Nearest Neighbor dan Support Vector Machine untuk menentukan prediksi produk-produk terlaris pada toko Madura memberikan tingkat ketepatan ramalan sehubungan dengan barang-barang yang laku terjual di toko tersebut [16]. Pemeriksaan ini menganalisis perhitungan K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM), ditemukan bahwa kedua perhitungan KNN dan SVM mencapai tingkat ketepatan yang sama pada angka 75%. Namun perhitungan SVM menghasilkan tingkat presisi yang lebih signifikan dibandingkan KNN. Selain itu penelitian lain terkait penerapan algoritma C4.5 untuk klasifikasi keberhasilan pengiriman barang yang bertujuan untuk menghasilkan data dalam pengelompokkan tingkat hasil pengangkutan barang dagangan [17], eksplorasi ini menerapkan prosedur penambahan informasi menggunakan strategi perhitungan C4.5 dan menggunakan pemrograman pendukung Rapid Miner mendapatkan tingkat presisi sebesar 93%. Confusion Matrix dan kurva ROC digunakan sebagai teknik penilaian untuk menghasilkan nilai AUC (Area Under Curve) yang memperoleh konsekuensi sebesar 0,739 dengan tingkat hasil (*Fair Characterization*).

Untuk menyikapi dan mengoptimalkan dinamika pasar yang terus berubah, penelitian ini berfokus pada aspek klasifikasi data penjualan makanan dan minuman online. Penelitian ini berfokus pada penerapan beberapa algoritma klasifikasi seperti SVM, KNN dan C5.0 untuk mengklasifikasikan penjualan makanan secara online, ditambah dengan pemanfaatan feature selection seperti penggunaan Chi Square dan Ensemble Tree Classifier dengan memilih 6 dan 10 fitur. Kombinasi ketiga algoritma dan feature selection diharapkan dapat memberikan hasil yang lebih akurat dan efisien dalam mengklasifikasikan jenis makanan dan permasalahan yang terjadi. Oleh karena itu, pemilik usaha dapat menggunakan hasil klasifikasi ini untuk membantu mengambil keputusan dalam menentukan strategi dalam pemasaran secara online [2], [6], [7].

2. METODE

Metode yang digunakan :

- a. Informasi demografis pelanggan terkait dengan usia pelanggan dan jenis kelamin.
- b. Melakukan pengecekan terkait dengan detail produk berdasarkan nama produk, kategori produk dan juga harga produk
- c. Menentukan informasi transaksi.
- d. Melakukan pengecekan terkait dengan data pengiriman.



Gambar 2 Tahapan Penelitian

Penelitian ini dimulai dengan mengidentifikasi masalah penjualan yang ada untuk memahami konteks dan fokus utama penelitian. Selanjutnya menentukan atribut yang relevan untuk memastikan kualitas data yang akan dianalisis. Data kemudian melalui tahap pre-processing yang mencakup pembersihan kolom, encoding untuk mengubah data kategori menjadi format numerik, dan normalisasi untuk menyamakan skala semua atribut. Selanjutnya, pemilihan fitur dilakukan menggunakan metode Chi Square dan Ensemble Tree Classifier untuk memilih 6 dan 10 fitur terbaik. Algoritma SVM, KNN, dan C4.5 digunakan untuk membangun model prediksi berdasarkan fitur-fitur yang telah dipilih. Model yang telah dibangun diuji untuk mengevaluasi kinerjanya, dan hasil penelitian disusun dalam bentuk laporan untuk dipublikasikan sehingga memberikan kontribusi ilmiah pada bidang terkait.

3. HASIL DAN PEMBAHASAN

PENGUMPULAN DATA

Dataset yang didapatkan merupakan kumpulan data mentah yang mencakup transaksi penjualan makanan secara online yang terdiri atas 48 kolom dan 1.885 baris. Terdapat beragam informasi mulai dari rincian pesanan hingga status pengiriman seperti yang ditampilkan pada gambar 2.

Gambar 3 Dataset awal

Berdasarkan data tersebut dapat diinformasikan terkait bisnis proses apa saja yang terdapat di dalamnya, diantaranya adalah:

1. Informasi Pesanan: Nomor pesanan, waktu pembuatan pesanan, dan perkiraan waktu pengiriman memberikan gambaran tentang dinamika pesanan.
2. Detail Produk: Dataset mencakup nama produk, SKU, variasi, harga awal, harga setelah diskon, dan total harga produk, memberikan wawasan mendalam tentang produk yang terjual.
3. Pembayaran: Rincian total pembayaran, diskon, dan metode pembayaran menyoroti aspek keuangan dari setiap transaksi.
4. Layanan Pengiriman: Pilihan antara antar ke counter atau pickup memberikan pemahaman tentang preferensi pelanggan terkait pengiriman.
5. Status Pesanan: Informasi apakah pesanan telah selesai atau dibatalkan memberikan indikasi kinerja layanan dan kepuasan pelanggan.

PRE-PROCESSING

Dataset yang ada sebelum digunakan pada pemodelan harus melalui serangkaian langkah preprocessing agar lebih efektif dan informatif, langkah-langkah yang dilakukan pada tahapan ini adalah sebagai berikut:

1. Pembersihan Kolom:

Menghapus atribut yang tidak memberikan kontribusi signifikan pada analisis sehingga akan menghasilkan dataset yang lebih terfokus dan mudah dipahami. Beberapa atribut yang dihapus diantaranya adalah source.name, nomor pesanan, alasan pembatalan, status pembatalan/pengembalian, no resi, pesanan harus dikirimkan sebelum, waktu pengiriman diatur, waktu pesanan dibuat, waktu pembayaran dilakukan, SKU induk, referensi SKU, nama variasi, total diskon, diskon dari penjual, diskon dari pembeli, diskon dari shopee, voucher di tanggung penjual, cashback coin, voucher di tanggung shopee, voucher diskon, paket diskon(shopee dan penjual), potongan coin shopee, diskon kartu kredit, ongkir pembeli, estimasi potongan biaya pengiriman, ongkir pengembalian barang, perkiraan ongkir, catatan pembeli, username, nama penerima, no.tlp, alamat pengiriman, dan waktu pesanan selesai.

2. Encoding Status Pesanan

Status pesanan direpresentasikan secara biner untuk memberi kemudahan pemahaman terkait dengan status pesanan, nilai 1 menunjukkan pesanan selesai dan nilai 0 menunjukkan pembatalan.

3. Encoding Opsi Layanan Antar

Opsi layanan antar direpresentasikan secara biner untuk mempermudah identifikasi cara pelanggan memilih penerimaan pesanan, di mana nilai 0 menandakan antar ke counter dan nilai 1 menunjukkan layanan *pickup*.

4. Normalisasi Harga dan Pembayaran

Harga awal, harga setelah diskon, total harga produk, dan total pembayaran dilakukan proses normalisasi menggunakan metode min-max.

5. Penambahan Label pada Produk dan Berat

Produk sekarang diberikan label yang jelas, sedangkan berat produk dinormalisasi. Langkah ini mendukung analisis yang lebih mendalam terkait preferensi produk dan karakteristik beratnya.

Dengan langkah-langkah preprocessing tersebut maka di dapatkan data yang siap untuk diolah seperti terlihat pada gambar 3.

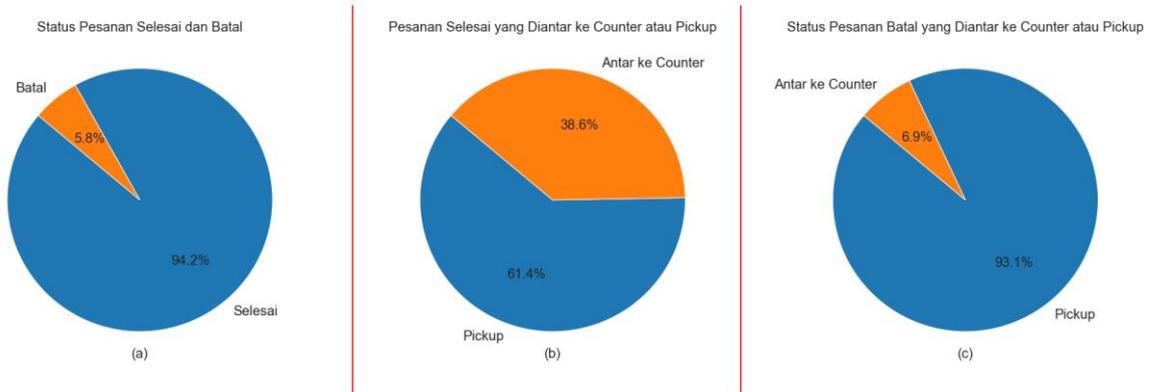
Jumlah	Total Harga Produk	Jumlah Produk di Pesan	Total Pembayaran	Kota/Kabupaten	Provinsi	Nama Produk	Berat Prod
1,0	0,0	2,0	0,0	KOTA TANGERANG	BANTEN	TABASYUK TABAS MINI PROZEN FREE CABAI RAWIT	400 GR
1,0	0,0	2,0	0,0	KOTA TANGERANG	BANTEN	TABASYUK MOLU (EMOJI DI KEJU)	200 GR
1,0	0,0	2,0	0,0	KAB. SIAMANGS	JAWA TENGAH	TABASYUK TABAS MINI PROZEN FREE CABAI RAWIT	400 GR
1,0	0,0	2,0	0,0	KAB. SIAMANGS	JAWA TENGAH	TABASYUK MOLU (EMOJI DI KEJU)	200 GR
1,0	0,0	2,0	0,0	KOTA JAKARTA PUKAT	DKI JAKARTA	TABASYUK TABAS MINI PROZEN FREE CABAI RAWIT	400 GR
1,0	0,0	2,0	0,0	KOTA JAKARTA PUKAT	DKI JAKARTA	TABASYUK MOLU (EMOJI DI KEJU)	200 GR
1,0	0,0	1,0	0,0	KOTA BEKASI	JAWA BARAT	TABASYUK DIMSUM AYAM LIDANS PROZEN FREE SAUS DIMSUM	300 GR
2,0	0,0	2,0	0,0	KAB. TANGERANG	BANTEN	MAKSER KNO3, KPSA, SPLY BLUNDELING PROMO	300 GR
1,0	0,0	1,0	0,0	KAB. CIAMIS	JAWA BARAT	TABASYUK TANGKUS JUMBO PROZEN FREE CABAI RAWIT	700 GR
2,0	0,0	2,0	0,0	KOTA JAKARTA TIMUR	DKI JAKARTA	TABASYUK TANGKUS JUMBO PROZEN FREE CABAI RAWIT	800 GR
1,0	0,0	1,0	0,0	KOTA TANGERANG	BANTEN	MAKSER KNO3, KPSA, SPLY BLUNDELING PROMO	300 GR
1,0	0,0	1,0	0,0	KOTA TANGERANG	BANTEN	MAKSER SPLY - 10 SPICE SPACE	150 GR
1,0	0,0	1,0	0,0	KOTA TANGERANG	BANTEN	MAKSER KNO3, KPSA, SPLY BLUNDELING PROMO	150 GR
1,0	0,0	1,0	0,0	KAB. CIAMIS	JAWA BARAT	TABASYUK TANGKUS JUMBO PROZEN FREE CABAI RAWIT	700 GR
1,0	0,1	10,0	0,1	KOTA TANGERANG SELATAN	BANTEN	TABASYUK TANGKUS JUMBO PROZEN FREE CABAI RAWIT	700 GR
1,0	0,0	6,0	0,0	KOTA DEPOK	JAWA BARAT	DANULI MABASA CIBADI DI KEJU ISI 10PCS	5750 GR
1,0	0,0	6,0	0,0	KOTA DEPOK	JAWA BARAT	DANULI MABASA SAMBAL AMBYAR	280 GR
2,0	0,0	10,0	0,1	KAB. BOGOR	JAWA BARAT	DANULI MABASA CIBADI DI KEJU ISI 10PCS	3020 GR
2,0	0,0	10,0	0,1	KAB. BOGOR	JAWA BARAT	DANULI MABASA CIBOKITEN	700 GR
2,0	0,0	10,0	0,1	KAB. BOGOR	JAWA BARAT	DANULI MABASA CIBUSUK	180 GR
2,0	0,0	10,0	0,1	KAB. BOGOR	JAWA BARAT	DANULI MABASA CIBUSUK	180 GR
2,0	0,0	10,0	0,1	KAB. BOGOR	JAWA BARAT	DANULI MABASA CIBUSUK	180 GR
1,0	0,0	1,0	0,0	KOTA DEPOK	JAWA BARAT	DANULI MABASA CIBADI DI KEJU ISI 10PCS	200 GR
2,0	0,0	2,0	0,0	KOTA BANDAR LAMPUNG	LAMPUNG	DANULI MABASA CIBADI DI KEJU ISI 10PCS	200 GR
2,0	0,0	2,0	0,0	KOTA DEPOK	JAWA BARAT	DANULI MABASA CIBADI DI KEJU ISI 10PCS	200 GR

Gambar 4 Data ternormalisasi

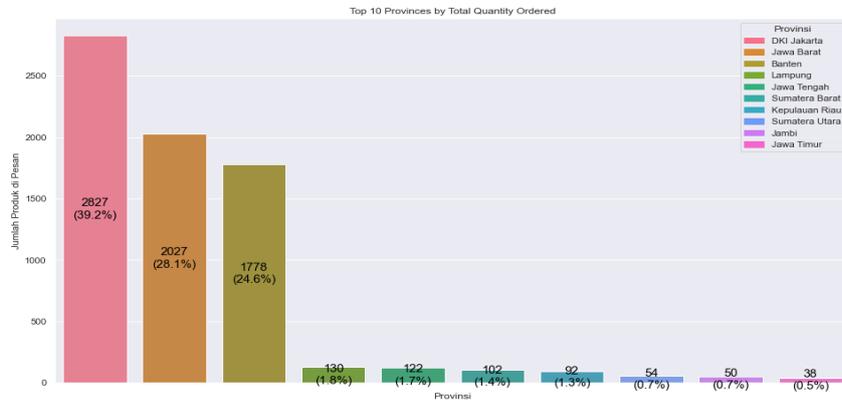
VISUALISASI

Setelah dilakukan pre-processing data dengan cermat, selanjutnya visualisasi dilakukan untuk mempermudah dalam menerima suatu informasi terkait dengan status pemesanan, pesanan yang diantar ke counter atau diambil, serta status pesanan yang batal diantar atau diambil, seperti ditampilkan pada gambar 4. Berdasarkan gambar 4(a) disimpulkan jumlah pesanan selesai sebesar 94,2% jauh lebih banyak jika dibandingkan dengan pesanan batal dengan nilai 5,8%. Sedangkan pada gambar 4(b) disimpulkan jumlah pesanan yang diambil sebesar 61,4%, jumlah ini lebih banyak jika dibandingkan dengan yang diantar ke counter sebesar 38,6%. Selanjutnya pada gambar 4(c) dapat disimpulkan jumlah pesanan yang diambil sebesar 93,1%, ini lebih banyak jika dibandingkan dengan yang diantar ke counter sebesar 6,9%. Berdasarkan gambar 5 dapat ditampilkan untuk 10 pesanan terbaik berdasarkan provinsi, ini dapat digunakan untuk mengetahui provinsi mana yang memang memberikan kontribusi terbesar untuk menghasilkan omset penjualan pada toko online ini, sehingga dapat dilakukan untuk persiapan produk apa saja yang laku berdasarkan provinsi. Terlihat tiga provinsi besar yang memiliki penjualan terbesar yaitu DKI Jakarta, Jawa Barat, dan Banten. Sehingga dari

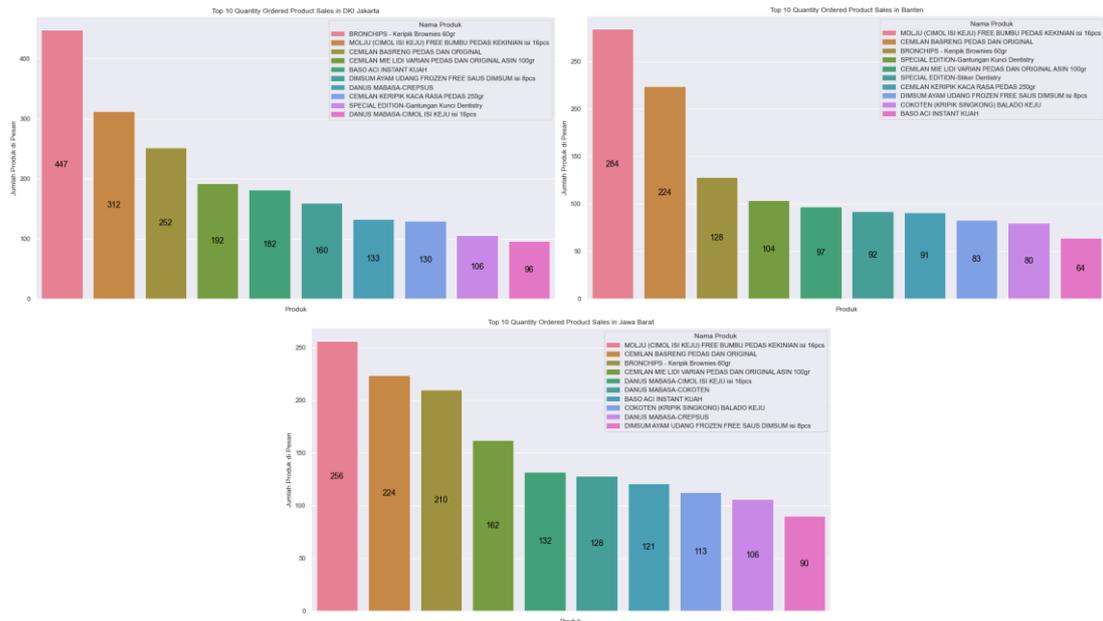
proses visualisasi dapat ditampilkan produk apa saja yang laris pada 3 provinsi tersebut seperti ditampilkan pada gambar 6.



Gambar 5 Grafik Pie (a) Status Pesanan Selesai dan Batal, (b) Pesanan Selesai yang Diantar ke Counter atau Pickup, (c) Status Pesanan Batal yang Diantar ke Counter atau Pickup



Gambar 6. 10 pesanan terbanyak berdasarkan provinsi

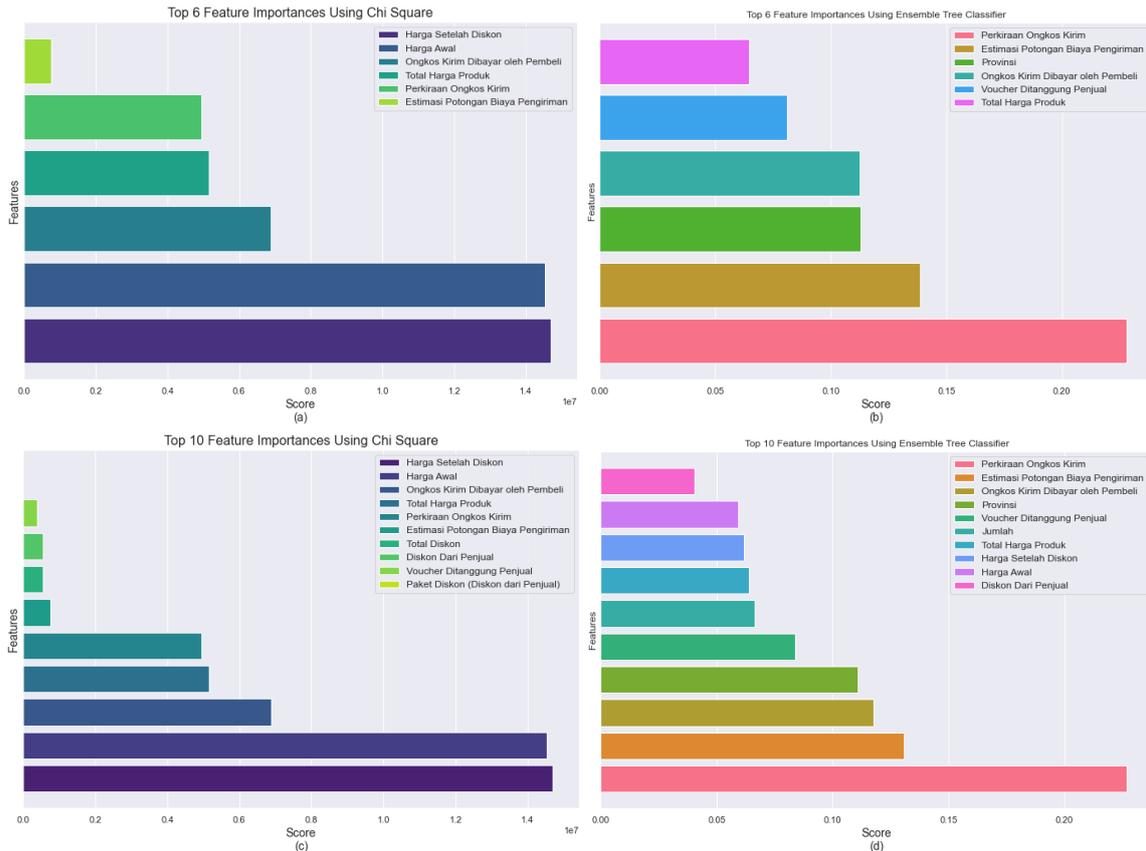


Gambar 7. Penjualan terbanyak berdasarkan 3 provinsi terbesar

Dapat disimpulkan dari gambar 6 bahwa dari ke-3 provinsi yang memiliki penjualan terbesar produk ini, produk yang sama-sama diminati dari ke-3 provinsi ini adalah Molju, cemilan basreng, broncips, dan juga cemilan mie lidi. Produk-produk tersebut sama-sama produk 5 besar pada masing-masing provinsi,

FEATURE SELECTION

Selanjutnya untuk mendapatkan hasil klasifikasi yang bagus, maka digunakan Chi Square untuk dilakukan pemilihan fitur yang digunakan. Dari sebanyak 48 kolom yang dimiliki oleh dataset ini, selanjutnya dengan pemanfaatan Chi Square peneliti menentukan 10 kolom dan 6 kolom, lalu menggunakan ensemble tree classifier untuk 10 kolom dan 6 kolom yang akan digunakan untuk proses klasifikasi, atribut-atribut yang memiliki peranan penting dalam klasifikasi ini dapat dilihat pada gambar 7(a), 7(b), 7(c), dan 7(d).



Gambar 8. Atribut terpilih menggunakan (a) chi square 6 fitur, (b) ensemble tree classifier 6 fitur, (c) chi square 10 fitur, (d) ensemble tree classifier 10 fitur

Berdasarkan gambar 7 baik menggunakan chi square maupun ensemble, fitur yang memiliki pengaruh adalah Perkiraan Ongkos Kirim, Ongkos Kirim Dibayar oleh Pembeli, Total Harga Produk, dan Estimasi Potongan Biaya Pengiriman. Dari ke-4 atribut yang paling berpengaruh tersebut, dapat disimpulkan masalah utama dari pembelian produk secara online adalah terkait dengan biaya pengiriman.

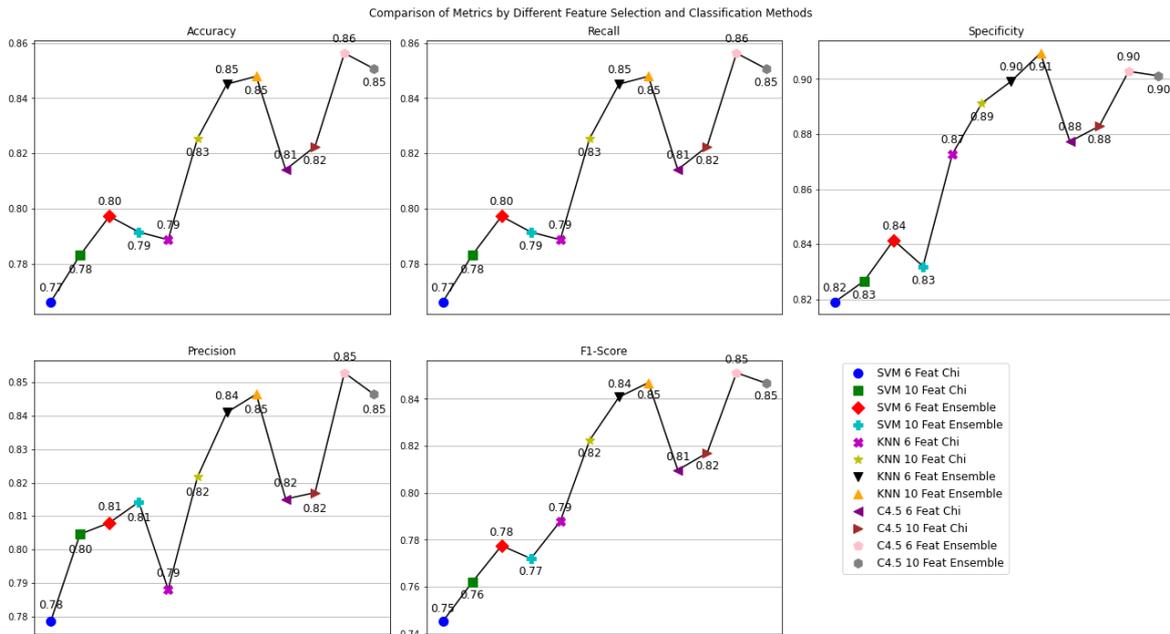
IMPLEMENTASI DAN PERBANDINGAN ALGORITMA

Untuk melakukan perbandingan berdasarkan penggunaan fitur yang memiliki kontribusi tinggi terhadap model yang digunakan, proses ini akan membagi data menjadi 2 bagian yaitu 80% untuk data latih dan 20% untuk data uji. Penggunaan chi square untuk model SVM akan menggunakan best hyperparameters untuk nilai C adalah 1, dan gamma='auto' dengan test accuracy best hyperparameters adalah 0,77. Sedangkan untuk ensemble tree classifier menggunakan best hyperparameters untuk nilai c adalah 10, dan gamma='auto' dengan test accuracy best hyperparameters adalah 0,80. Pada pemodelan menggunakan KNN, seluruh data yang digunakan dalam proses sama-sama akan menggunakan nilai k=1 berdasarkan pemanfaatan cross validation. Sedangkan pada penggunaan model C4.5 hasil best hyperparameters untuk seluruh data yang digunakan dalam proses pemodelan sama-sama menggunakan nilai 'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 2. Seluruh pemanfaatan best hyperparameters dapat dilihat pada tabel 1.

Tabel 1. Hasil evaluasi model klasifikasi dan pemanfaatan feature selection

Pemodelan	SVM				KNN				C4.5			
	Chi		Ensemble Tree Classifier		Chi		Ensemble Tree Classifier		Chi		Ensemble Tree Classifier	
	6 Feat	10 Feat	6 Feat	10 Feat	6 Feat	10 Feat	6 Feat	10 Feat	6 Feat	10 Feat	6 Feat	10 Feat
Gamma	auto	auto	auto	auto								
Nilai C	1	1	10	10								
Nilai k					1	1	1	1				
Max_depth									10	10	10	10
Min samples leaf									1	1	1	1
Mean samples split									2	2	2	2
test accuracy best	0,77	0,78	0,80	0,80	0,79	0,83	0,85	0,85	0,81	0,82	0,85	0,85
hyperparameters												
Accuracy	76,62	78,31	79,72	79,15	78,87	82,54	84,51	84,79	81,41	82,25	85,63	85,07
Recall	76,62	78,31	79,72	79,15	78,87	82,54	84,51	84,79	81,41	82,25	85,63	85,07
Specificity	81,91	82,67	84,15	83,20	87,28	89,12	89,92	90,94	87,73	88,29	90,28	90,12
Precision	77,85	80,46	80,80	81,42	78,81	82,17	84,10	84,64	81,51	81,70	85,28	84,64
F1-Score	74,52	76,19	77,75	77,20	78,79	82,24	84,08	84,69	80,97	81,70	85,11	84,67

Hasil dari nilai evaluasi pada masing-masing model agar dapat dengan mudah terlihat model klasifikasi mana yang memiliki nilai akurasi, presisi, recall, dan F1-score tertinggi atau terendah dapat dilihat berdasarkan visualisasi pada gambar 8



Gambar 9. Hasil evaluasi pemodelan 80:20

Berdasarkan gambar 8 telah mendapatkan hasil evaluasi dari masing-masing model, dan dinyatakan bahwa model C4.5 dengan menggunakan data hasil feature selection menggunakan ensemble tree classifier 6 fitur memiliki nilai akurasi tertinggi sebesar 0,86%, pun demikian untuk nilai recall, precision, dan juga F1-score yang masing-masing bernilai sebesar 0,86%, 0,85%, dan 0,85%. Selanjutnya nilai terendah dari pemodelan tersebut menggunakan SVM dengan penggunaan feature selection chi square dengan nilai akurasi, recall, presisi, dan F1-score masing-masing sebesar 0,77%, 0,77%, 0,78%, dan 0,75%.

4. PENUTUP

Penelitian ini digunakan untuk membantu UMKM dalam melakukan penjualan yang dilakukan secara online. Berdasarkan hasil feature selection yang menggunakan metode chi square dan juga ensemble tree classifier telah didapatkan bahwa faktor yang paling berpengaruh terhadap data tersebut terkait dengan biaya pengiriman yang dilakukan oleh pembeli. Kolom-kolom tersebut yang berpengaruh terkait dengan Perkiraan Ongkos Kirim, Ongkos Kirim Dibayar oleh Pembeli, Total Harga Produk, dan Estimasi Potongan Biaya Pengiriman. Selanjutnya dilihat dari sebaran pembelian berdasarkan provinsi, paling besar transaksi dilakukan berasal dari provinsi yang dekat dengan pelaku UMKM, yaitu berada di provinsi DKI Jakarta, Jawa Barat, dan Banten. Provinsi tersebut sangat berhubungan sekali dengan biaya pengiriman, karena semakin dekat jarak penjual dengan pembeli, maka akan semakin murah juga biaya pengiriman yang harus dikeluarkan oleh pembeli. Selanjutnya setelah dilakukan evaluasi menggunakan 3 model algoritma yang digunakan, yaitu SVM, KNN, dan C4.5 menyebutkan bahwa nilai akurasi tertinggi diperoleh saat menggunakan model C4.5 dengan data yang digunakan berasal dari ensemble tree classifier dengan 6 fitur sebesar 0.86%, lalu diikuti oleh model C4.5 10 fitur, KNN 6 fitur, dan KNN 10 fitur yang semua sumber data berasal dari ensemble tree classifier dengan masing-masing nilai akurasi sebesar 0,85%.

UCAPAN TERIMAKASIH

Terima kasih pada Universitas Mercu Buana melalui LPPM yang telah mendanai kegiatan penelitian internal skema dosen muda ini.

DAFTAR PUSTAKA

- [1] K. Indonesia, "UMKM Indonesia," 2024. <https://kadin.id/data-dan-statistik/umkm-indonesia/>.
- [2] P. L. G. Canggih Gumanky Farunik, "Tantangan dan Peluang Bisnis dalam Beradaptasi dengan Pasar Generasi Z," *J. Ekon. dan Biana*, vol. 3, no. 1, pp. 487-496., 2023.
- [3] "Daftar UMKM Indonesia," *KEMENKOPUKM*, 2024. <https://umkm.depkop.go.id/>.
- [4] A. Yuliani, "Kemenkop UKM: 3,79 Juta UMKM Sudah Go Online," *KOMINFO*. https://www.kominfo.go.id/content/detail/11526/kemenkop-ukm-379-juta-umkm-sudah-go-online/0/sorotan_media.
- [5] UMKM, "Peningkatan Jumlah UMKM Tangerang yang Go Digital," *GETI Media*, 2023. <https://getimedia.id/2023/12/13/peningkatan-jumlah-umkm-tangerang-yang-go-digital/>.
- [6] R. Putriana and S. Kholil, "Analisis Trik dan Teknik Marketing Generasi Z dalam Mempromosikan Produk Somethinc di Instagram," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 2, pp. 1676–1684, 2024, [Online]. Available: <https://journal.stmiki.ac.id>.
- [7] H. Wijoyo, I. Indrawan, Y. Cahyono, A. L. Handoko, and R. Santamoko, *Generasi Z Dan Revolusi Industri 4.0*, vol. 1. 2020.
- [8] D. A. Mukhsinin, M. Rafliansyah, S. A. Ibrahim, R. Rahmaddeni, and D. Wulandari, "Implementasi Algoritma Decision Tree untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 570–579, 2024, doi: 10.57152/malcom.v4i2.1255.
- [9] R. N. Ramadhon, A. Ogi, A. P. Agung, R. Putra, S. S. Febrihartina, and U. Firdaus, "Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank," *Karimah Tauhid*, vol. 3, no. 2, pp. 1860–1874, 2024, doi: 10.30997/karimahtauhid.v3i2.11952.
- [10] M. Arif Saifudin, H. Endah Wahanani, and A. Junaidi, "Implementasi Algoritma Asosiasi Fp-Growth Dan Klasifikasi K-Means Terhadap Pola Pembelian Konsumen Di Marketplace Shopee," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 764–771, 2024, doi: 10.36040/jati.v8i1.8848.
- [11] B. Harmadi, P. S. Matematika, and F. Sains dan Teknologi, "Penerapan Algoritma Decision Tree Dalam Segmentasi Customer," vol. 12, no. 2, pp. 315–322, 2022.
- [12] A. P. A. S. M R, J. Sodik, M. Saripudin, S. M. Rejeki, and U. Firdaus, "Kajian untuk Segmentasi Customer Bank dengan Algoritma K-Means," *Karimah Tauhid*, vol. 3, no. 3, pp. 3899–3906, 2024, doi: 10.30997/karimahtauhid.v3i3.12318.
- [13] R. Novalana and wasino, "Pengaruh Persepsi Harga, Kualitas Pelayanan Dan Lokasi Terhadap Keputusan Pembelian Di Leora Kopi," *J. Panor. Nusant.*, vol. Vol.18, no. Issue. 2, pp. 1–11, 2023, [Online]. Available: <https://ejournal.asaindo.ac.id/index.php/panoramanusantara/issue/view/185>.
- [14] T. Tukino, "Analisis Prediksi Penjualan Sayuran Hidroponik Menggunakan Algoritma C4.5 pada PT Batam Indo Agri Perkasa," *J. Desain Dan Anal. Teknol.*, vol. 2, no. 2, pp. 129–137, 2023, doi: 10.58520/jddat.v2i2.30.
- [15] F. Firmansyah and O. Nurdiawan, "Penerapan Data Mining Menggunakan Algoritma Frequent Pattern - Growth Untuk Menentukan Pola Pembelian Produk Chemicals," *JATI (Jurnal Mhs. Tek. Inform.)*, vol.

- 7, no. 1, pp. 547–551, 2023, doi: 10.36040/jati.v7i1.6371.
- [16] I. H. Pratama and U. Salamah, “Perbandingan Algoritma K-Nearest Neighbor Dan Support Vector Machine Untuk Menentukan Prediksi Produk-Produk Terlaris Pada Toko Madura Kecamatan Pondok Aren,” *J. Tek. Inform. Kaputama*, vol. 6, no. 2, pp. 846–858, 2022, [Online]. Available: <https://jurnal-backup.kaputama.ac.id/index.php/JTIK/article/view/1163>.
- [17] G. Taufik and D. Jatmika, “Penerapan Algoritma C4 . 5 Untuk Klasifikasi Keberhasilan Pengiriman Barang,” pp. 12–26, 2021.