



Penerapan Data Mining Untuk Prediksi Penjualan Roti Terlaris Menggunakan Metode K-Nearest Neighbor

Arifin Munthe*, Kusmanto, Basyarul Ulya

Universitas Al Washliyah Labuhanbatu, Sumatera Utara, Indonesia

Email: ^{1*}arifinmunthe1307@gmail.com, ²kusnabara03@gmail.com, ³ulyabasyarul@gmail.com

(* :coressponding: arifinmunthe1307@gmail.com)

Abstrak- Penelitian ini dilakukan untuk memprediksi penjualan roti terlaris menggunakan metode K-Nearest Neighbor (KNN) pada Pabrik Roti SAMAWA Rantauprapat. Dengan memanfaatkan teknik data mining, penelitian ini bertujuan memberikan informasi yang akurat bagi pengambilan keputusan penjualan. Data mining merupakan proses untuk menemukan pola tersembunyi dalam data, sedangkan metode KNN adalah algoritma klasifikasi yang bekerja berdasarkan kedekatan jarak antar data. Model klasifikasi ini efektif untuk menganalisis data numerik dan menetapkan data baru berdasarkan data historis. Penelitian ini menggunakan pendekatan kuantitatif dengan tahapan pengumpulan data, preprocessing, pembagian data training dan testing, serta evaluasi model menggunakan aplikasi Orange. Data yang digunakan sebanyak 150 data penjualan, dan analisis dilakukan dengan menggunakan algoritma KNN untuk proses klasifikasi. Hasil klasifikasi menunjukkan bahwa dari 100 data testing, sebanyak 59 diklasifikasikan sebagai "laris" dan 41 sebagai "tidak laris". Evaluasi model menghasilkan nilai akurasi sebesar 85% dan AUC sebesar 0,948, yang menunjukkan performa model cukup tinggi. Berdasarkan hasil tersebut, dapat disimpulkan bahwa metode KNN efektif digunakan dalam memprediksi penjualan roti terlaris. Model ini diharapkan dapat membantu pihak pabrik dalam mengelola stok dan strategi pemasaran dengan lebih baik.

Kata Kunci: *Data Mining, K-Nearest Neighbor, Prediksi Penjualan, Roti Terlaris, Model Klasifikasi*

Abstract- This research was conducted to predict the best-selling bread sales using the K-Nearest Neighbor (KNN) method at SAMAWA Rantauprapat Bread Factory. By utilizing data mining techniques, this research aims to provide accurate information for sales decision making. Data mining is the process of finding hidden patterns in data, while the KNN method is a classification algorithm that works based on the closeness of the distance between data. This classification model is effective for analyzing numerical data and mapping new data based on historical data. This research uses a quantitative approach with the stages of data collection, preprocessing, division of training and testing data, and model evaluation using the Orange application. The data used was 150 sales data, and the analysis was carried out using the KNN algorithm for the classification process. The classification results showed that out of 100 testing data, 59 were classified as "in demand" and 41 as "not in demand". The evaluation of the model resulted in an accuracy value of 85% and an AUC of 0.948, which shows that the performance of the model is quite high. Based on these results, it can be concluded that the KNN method is effective in predicting best-selling bread sales. This model is expected to help the factory in managing stock and marketing strategies better.

Keywords: *Data Mining, K-Nearest Neighbor, Sales Prediction, Best Selling Bread, Classification Model*

1. PENDAHULUAN

Perkembangan teknologi informasi yang semakin pesat telah membawa perubahan signifikan dalam berbagai sektor, termasuk di dalamnya sektor industri makanan. Industri ini tidak hanya dituntut untuk memproduksi barang dalam jumlah besar, tetapi juga dituntut untuk memahami dan memprediksi perilaku pasar agar dapat bersaing secara efektif. Salah satu subsektor industri makanan yang mengalami pertumbuhan cukup stabil adalah industri roti, yang memiliki permintaan tinggi dari berbagai lapisan masyarakat. Produk roti, dengan berbagai variasi rasa dan bentuk, menjadi konsumsi harian yang praktis dan digemari oleh semua kalangan. Dalam hal ini, produsen roti dituntut untuk selalu siap dalam memenuhi kebutuhan pasar yang dinamis dan terkadang sulit diprediksi. Untuk menjawab tantangan tersebut, pemanfaatan teknologi berbasis data menjadi sangat penting dalam mendukung keputusan strategis, khususnya dalam aspek produksi dan distribusi.

Pabrik Roti SAMAWA di Rantauprapat sebagai salah satu pelaku industri roti skala menengah, menghadapi tantangan dalam mengelola produksi dan stok barang secara efektif. Seringkali, ketidaksesuaian antara jumlah roti yang diproduksi dengan kebutuhan pasar menyebabkan terjadinya penumpukan stok atau sebaliknya, kekurangan barang. Kedua kondisi tersebut tentu merugikan pihak produsen, baik dari sisi biaya produksi maupun dari sisi kepuasan pelanggan. Oleh karena itu, kemampuan untuk memprediksi produk mana yang berpotensi laris dalam periode tertentu menjadi kebutuhan mendesak. Dalam konteks inilah, pemanfaatan metode analisis data seperti data mining menjadi sangat relevan. Data mining dapat membantu menemukan pola tersembunyi dalam data historis penjualan yang sebelumnya tidak terdeteksi secara kasat mata, sehingga dapat digunakan untuk mendukung pengambilan keputusan yang lebih tepat dan efisien.

Namun dalam praktiknya, proses prediksi penjualan di Pabrik Roti SAMAWA masih dilakukan secara manual atau berdasarkan intuisi manajerial yang terkadang tidak didukung oleh data historis yang akurat. Pendekatan semacam ini sangat rentan terhadap kesalahan, terlebih ketika volume data penjualan semakin meningkat dan kompleksitas perilaku konsumen juga berubah. Akibatnya, penentuan jumlah produksi harian dan pengelolaan stok sering tidak tepat sasaran, sehingga menimbulkan inefisiensi operasional. Selain itu, metode manual tidak mampu memberikan hasil prediksi yang konsisten, dan sulit untuk dijadikan dasar dalam evaluasi dan perencanaan jangka panjang. Oleh sebab itu, dibutuhkan



sebuah pendekatan yang lebih sistematis dan berbasis data untuk meningkatkan akurasi prediksi penjualan serta mengoptimalkan proses produksi.

Meskipun telah banyak penelitian terkait prediksi penjualan menggunakan berbagai metode data mining, seperti algoritma Decision Tree, Naïve Bayes, dan Artificial Neural Network, namun penelitian yang secara spesifik menerapkan metode K-Nearest Neighbor (KNN) dalam konteks prediksi penjualan roti masih tergolong minim, khususnya dalam skala industri kecil-menengah seperti Pabrik Roti SAMAWA. Kebanyakan penelitian lebih difokuskan pada industri besar atau sektor lain seperti ritel dan e-commerce. Padahal, metode KNN memiliki keunggulan dalam hal kesederhanaan dan kemampuannya dalam melakukan klasifikasi berdasarkan kemiripan data. Dengan kata lain, KNN sangat cocok digunakan pada kasus-kasus dengan data historis yang tidak terlalu besar namun memiliki pola yang jelas, seperti pada penjualan harian roti. Gap inilah yang menjadi dasar penting bagi penelitian ini untuk mengkaji efektivitas penerapan metode KNN dalam memprediksi penjualan roti terlaris.

Sebagai respons terhadap permasalahan dan gap yang telah diidentifikasi, penelitian ini bertujuan untuk merancang dan mengimplementasikan model klasifikasi berbasis metode K-Nearest Neighbor (KNN) yang mampu memprediksi produk roti mana yang berpotensi mengalami penjualan tinggi. Dengan memanfaatkan data historis penjualan yang dimiliki oleh Pabrik Roti SAMAWA, model ini diharapkan mampu mengenali pola-pola penjualan dari berbagai jenis roti berdasarkan atribut-atribut tertentu seperti harga, stok awal, hari penjualan, dan hasil penjualan sebelumnya. Model ini akan dirancang menggunakan tools visualisasi dan pemodelan data seperti Orange Data Mining yang memudahkan proses perancangan dan evaluasi model klasifikasi. Dengan pendekatan ini, proses pengambilan keputusan tidak lagi bergantung pada intuisi semata, tetapi berdasarkan hasil analisis data yang objektif dan dapat dipertanggungjawabkan.

Melalui penerapan metode K-Nearest Neighbor dalam penelitian ini, diharapkan Pabrik Roti SAMAWA dapat memperoleh sistem pendukung keputusan yang mampu memprediksi produk-produk roti yang berpotensi laris dengan akurasi yang tinggi. Hasil dari klasifikasi ini dapat digunakan untuk menentukan jumlah produksi yang tepat setiap harinya, mengurangi risiko penumpukan stok, serta meningkatkan kepuasan pelanggan melalui ketersediaan produk yang sesuai permintaan pasar. Lebih jauh lagi, penelitian ini diharapkan dapat menjadi acuan atau rujukan bagi industri kecil-menengah lainnya yang ingin menerapkan teknologi data mining dalam kegiatan operasionalnya. Dengan demikian, penerapan teknologi berbasis data tidak hanya menjadi bagian dari transformasi digital di perusahaan besar, tetapi juga dapat diadopsi secara efektif oleh pelaku industri lokal sebagai bentuk adaptasi terhadap era industry modern.

2. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah metode kuantitatif dengan pendekatan data mining menggunakan algoritma K-Nearest Neighbor (KNN) untuk melakukan klasifikasi terhadap data penjualan roti [1]. Penelitian ini mengikuti tahapan umum dalam proses data mining, yang mencakup pengumpulan data, preprocessing data, pemisahan data menjadi data training dan data testing, perancangan model klasifikasi, hingga evaluasi model [2] [3]. Data yang digunakan terdiri dari 150 data penjualan yang memuat beberapa variabel seperti nama produk roti, harga, stok awal, penjualan hari sebelumnya, dan hari transaksi. Penelitian ini menggunakan aplikasi Orange sebagai alat bantu dalam proses klasifikasi dan evaluasi model [4].

Penelitian ini dilakukan di Pabrik Roti SAMAWA yang berlokasi di Rantauprapat, Kabupaten Labuhan Batu, Sumatera Utara. Lokasi ini dipilih karena Pabrik Roti SAMAWA merupakan salah satu produsen roti lokal yang memiliki jumlah transaksi penjualan harian yang cukup tinggi dan beragam jenis produk roti. Oleh karena itu, data yang tersedia dinilai representatif untuk dianalisis guna mengetahui produk roti mana yang termasuk dalam kategori “laris” maupun “tidak laris”. Dengan memanfaatkan metode KNN, diharapkan penelitian ini dapat membantu pihak pabrik dalam mengambil keputusan strategis terkait stok dan promosi berdasarkan prediksi penjualan roti.

2.1 Data Mining

Data mining merupakan proses penggalian informasi atau pengetahuan yang sebelumnya tidak diketahui dari sejumlah besar data yang tersimpan dalam basis data [5] [6]. Proses ini melibatkan berbagai teknik statistik, matematika, dan kecerdasan buatan untuk mengidentifikasi pola-pola tersembunyi, tren, dan hubungan yang bermakna di dalam data. Dalam konteks bisnis, data mining sangat bermanfaat untuk pengambilan keputusan yang lebih tepat karena dapat memberikan prediksi atau klasifikasi berdasarkan data historis [7]. Beberapa proses utama dalam data mining meliputi seleksi data, transformasi data, pemodelan, evaluasi, dan penerapan hasil [8]. Dengan demikian, data mining menjadi alat penting dalam mengelola informasi dan mendukung strategi organisasi yang berbasis data.

2.2 Metode K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) adalah salah satu algoritma klasifikasi yang bersifat non-parametrik dan sederhana namun cukup efektif dalam mengolah data numerik [9] [10]. Metode ini bekerja dengan cara membandingkan jarak antara data baru yang ingin diklasifikasikan dengan sejumlah data lain yang telah memiliki label kelas, lalu menentukan kelas berdasarkan mayoritas dari data tetangga terdekat (nearest neighbors) [11]. Jarak yang digunakan dalam KNN biasanya dihitung menggunakan Euclidean Distance, meskipun dapat pula menggunakan metode jarak lain tergantung pada kebutuhan [12]. Keunggulan dari metode KNN adalah kemampuannya dalam bekerja dengan data yang memiliki pola



kedekatan serta kemudahan implementasinya, meskipun kekurangannya terletak pada efisiensi ketika diterapkan pada data berukuran besar [13].

2.3 Model Klasifikasi

Model klasifikasi adalah suatu metode dalam data mining yang digunakan untuk memetakan data ke dalam kategori atau kelas tertentu berdasarkan atribut atau variabel input yang dimilikinya [14] [15] [16]. Model ini dilatih menggunakan data historis (data training) yang telah diketahui kelasnya, sehingga mampu mempelajari pola-pola tertentu untuk kemudian digunakan dalam mengklasifikasikan data baru [17]. Proses klasifikasi sangat penting dalam berbagai aplikasi seperti deteksi penipuan, analisis kredit, prediksi pasar, dan tentu saja, prediksi penjualan seperti dalam penelitian ini [18]. Dalam pembangunannya, model klasifikasi dapat menggunakan berbagai algoritma, termasuk KNN, Decision Tree, Naïve Bayes, dan lain sebagainya, tergantung pada jenis data serta kebutuhan analisis.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan pendekatan simulasi berdasarkan kondisi nyata di Pabrik Roti SAMAWA Rantaupratat. Total data yang digunakan sebanyak 150 data transaksi yang mencakup berbagai jenis roti. Variabel yang digunakan dalam penelitian ini terdiri dari Nama Produk Roti, Harga Jual, Stok Awal, Penjualan Hari Sebelumnya, dan Hari Transaksi. Selain itu, variabel target atau variabel dependen yang dianalisis adalah Status Penjualan yang diklasifikasikan menjadi Laris atau Tidak Laris berdasarkan persentase penjualan terhadap stok awal. Data ini selanjutnya digunakan untuk membangun model prediksi menggunakan algoritma K-Nearest Neighbor (KNN).

Tabel 1. Data Mentah

No	Nama Produk Roti	Harga Jual (Rp)	Stok Awal	Penjualan Hari Sebelumnya	Hari	Status
1.	Roti Tawar	Rp 5,000.00	34	10	Sabtu	Tidak Laris
2.	Roti Coklat	Rp 5,000.00	31	14	Jumat	Tidak Laris
3.	Roti Keju	Rp 5,000.00	49	46	Sabtu	Laris
4.	Roti Coklat	Rp 5,000.00	38	21	Rabu	Tidak Laris
5.	Roti Tawar	Rp 5,000.00	31	21	Minggu	Tidak Laris
6.	Roti Tawar	Rp 5,000.00	41	28	Jumat	Tidak Laris
7.	Roti Coklat	Rp 5,000.00	43	22	Selasa	Tidak Laris
8.	Roti Keju	Rp 5,000.00	49	15	Rabu	Tidak Laris
9.	Roti Coklat	Rp 5,000.00	35	30	Minggu	Laris
10.	Roti Keju	Rp 5,000.00	46	41	Jumat	Laris
.. 150	.. Roti Keju	Rp .. 5,000.00	.. 35	.. 26	.. Rabu	.. Laris

Pada tabel di atas merupakan data yang sudah diperoleh untuk penelitian ini, dengan jumlah total 150 data yang mencakup informasi transaksi penjualan roti. Data tersebut mencakup berbagai variabel seperti harga jual, stok awal, penjualan hari sebelumnya, hari transaksi, dan status penjualan (Laris atau Tidak Laris), yang akan digunakan untuk menganalisis pola penjualan dan membangun model prediksi menggunakan metode K-Nearest Neighbor (KNN).

3.2 Preprocessing Data

Pada tahap preprocessing data, dilakukan proses pembersihan dan seleksi untuk memastikan bahwa data yang digunakan dalam penelitian ini benar-benar valid dan layak untuk dianalisis. Pembersihan data mencakup pengecekan terhadap data yang tidak konsisten, duplikat, atau data kosong yang dapat mempengaruhi akurasi model. Setelah data dinyatakan bersih, dilakukan pemisahan dataset menjadi dua bagian, yaitu data training dan data testing. Data training digunakan untuk membangun model prediksi, sementara data testing digunakan untuk menguji dan mengevaluasi performa model yang telah dibangun.



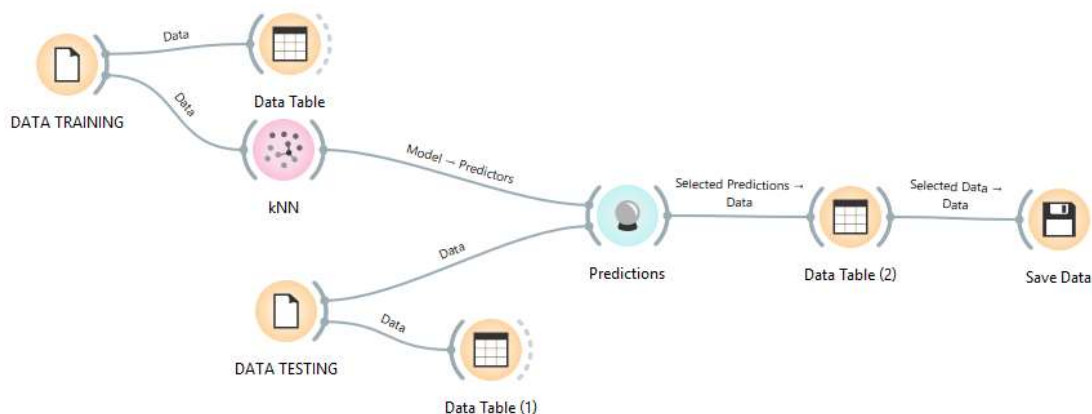
Tabel 2. Data Training

No	Nama Produk Roti	Harga Jual (Rp)	Stok Awal	Penjualan Hari Sebelumnya	Hari	Status
1.	Roti Tawar	Rp 5,000.00	34	10	Sabtu	Tidak Laris
2.	Roti Coklat	Rp 5,000.00	31	14	Jumat	Tidak Laris
3.	Roti Keju	Rp 5,000.00	49	46	Sabtu	Laris
4.	Roti Coklat	Rp 5,000.00	38	21	Rabu	Tidak Laris
5.	Roti Tawar	Rp 5,000.00	31	21	Minggu	Tidak Laris
6.	Roti Tawar	Rp 5,000.00	41	28	Jumat	Tidak Laris
7.	Roti Coklat	Rp 5,000.00	43	22	Selasa	Tidak Laris
8.	Roti Keju	Rp 5,000.00	49	15	Rabu	Tidak Laris
9.	Roti Coklat	Rp 5,000.00	35	30	Minggu	Laris
10.	Roti Keju	Rp 5,000.00	46	41	Jumat	Laris
....
50	Roti Tawar	Rp 5,000.00	45	27	Jumat	Tidak Laris

Pada tabel di atas merupakan data training penelitian yang digunakan sebanyak 50 data, yang berfungsi untuk melatih model algoritma K-Nearest Neighbor (KNN) dalam mengenali pola-pola yang terdapat pada data. Data training ini menjadi dasar bagi model untuk membangun pemahaman hubungan antara variabel-variabel independen seperti Nama Produk Roti, Harga Jual, Stok Awal, Penjualan Hari Sebelumnya, dan Hari Transaksi terhadap variabel target, yaitu Status Penjualan (Laris atau Tidak Laris).

Perancangan Model Klasifikasi

Pada tahap perancangan model klasifikasi, dilakukan proses untuk membangun model yang akan digunakan dalam penelitian ini. Model ini dirancang berdasarkan algoritma K-Nearest Neighbor (KNN) yang bertujuan untuk melakukan klasifikasi terhadap data sampel yang tersedia. Pada tahapan ini, model dikembangkan dengan memperhatikan variabel-variabel yang telah dipilih, sehingga mampu mengenali pola dan hubungan antar data. Model yang telah dirancang ini nantinya akan digunakan untuk mengklasifikasikan status penjualan roti menjadi kategori Laris atau Tidak Laris berdasarkan data training dan data testing yang telah disiapkan.


Gambar 1. Perancangan Model Klasifikasi

Pada gambar di atas merupakan model yang akan digunakan dalam penelitian ini. Model tersebut dirancang dengan menggunakan metode K-Nearest Neighbor (KNN) dan dibangun menggunakan aplikasi Orange. Pada model yang ditampilkan, metode KNN dapat dikenali melalui node berwarna merah yang menunjukkan algoritma klasifikasi yang digunakan. Model ini akan digunakan untuk melakukan proses klasifikasi terhadap data sampel guna memprediksi status penjualan roti menjadi kategori Laris atau Tidak Laris.

Hasil Klasifikasi

Pada hasil klasifikasi, output dari model akan dipaparkan dalam bentuk tabel untuk memudahkan analisis dan interpretasi. Tabel tersebut akan menampilkan hasil prediksi status penjualan roti berdasarkan data testing, sehingga dapat diketahui apakah roti termasuk dalam kategori Laris atau Tidak Laris sesuai dengan klasifikasi yang dilakukan oleh model K-Nearest Neighbor (KNN).

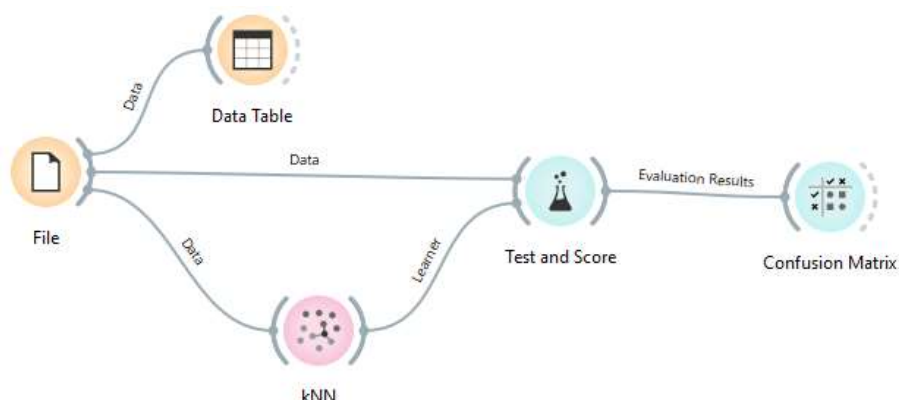
Tabel 3. Hasil Klasifikasi

Nama Produk Roti	Harga Jual (Rp)	Stok Awal	Penjualan Hari Sebelumnya	Hari	Status
Roti Sobek	Rp 5,000.00	37	29	Rabu	Laris
Roti Manis	Rp 5,000.00	21	11	Minggu	Tidak Laris
Roti Coklat	Rp 5,000.00	22	20	Selasa	Laris
Roti Keju	Rp 5,000.00	26	26	Selasa	Laris
Roti Coklat	Rp 5,000.00	24	21	Senin	Laris
Roti Tawar	Rp 5,000.00	42	13	Rabu	Tidak Laris
Roti Tawar	Rp 5,000.00	26	10	Selasa	Tidak Laris
Roti Sobek	Rp 5,000.00	32	11	Minggu	Laris
Roti Coklat	Rp 5,000.00	22	20	Selasa	Laris
Roti Sobek	Rp 5,000.00	60	14	Minggu	Laris

Pada tabel di atas, tercantum hasil klasifikasi yang telah dilakukan menggunakan model klasifikasi yang telah dirancang sebelumnya dengan metode K-Nearest Neighbor (KNN). Hasil klasifikasi ini diperoleh dari 100 data sampel yang digunakan dalam penelitian. Berdasarkan analisis yang dilakukan, sebanyak 59 data berhasil dikategorikan dalam status Laris, sementara 41 data lainnya terklasifikasi sebagai Tidak Laris. Hasil ini menunjukkan bahwa model klasifikasi yang diterapkan memiliki kemampuan untuk memprediksi status penjualan dengan proporsi yang cukup signifikan, meskipun terdapat sejumlah data yang tidak terklasifikasi sebagai laris. Evaluasi lebih lanjut dapat dilakukan untuk meningkatkan akurasi model dalam klasifikasi lebih lanjut.

Perancangan Model Evaluasi

Pada tahap perancangan model evaluasi, dilakukan proses penyusunan model yang akan digunakan untuk mengukur dan menilai kinerja metode klasifikasi yang diterapkan dalam penelitian ini. Model evaluasi dirancang untuk menguji tingkat akurasi, presisi, recall, dan f1-score dari metode K-Nearest Neighbor (KNN) yang digunakan. Dengan adanya perancangan model evaluasi ini, diharapkan dapat diperoleh gambaran yang objektif mengenai sejauh mana metode yang dipilih mampu mengklasifikasikan data dengan baik berdasarkan data testing yang telah disiapkan.

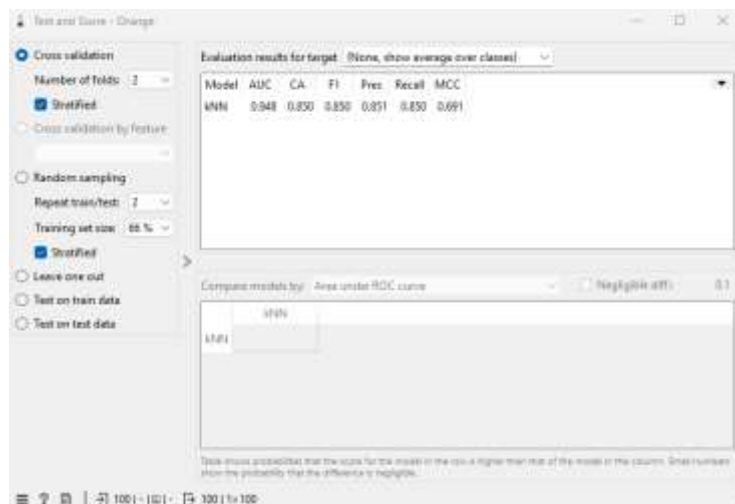


Gambar 2. Perancangan Model Evaluasi

Pada gambar di atas merupakan model evaluasi yang digunakan untuk menguji kemampuan metode dalam melakukan klasifikasi data. Model evaluasi ini dirancang untuk mengukur performa algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan data sampel dengan benar. Melalui model ini, hasil klasifikasi akan dievaluasi menggunakan metrik seperti akurasi, presisi, recall, dan f1-score, sehingga dapat diketahui sejauh mana tingkat keberhasilan metode dalam memprediksi status penjualan roti.

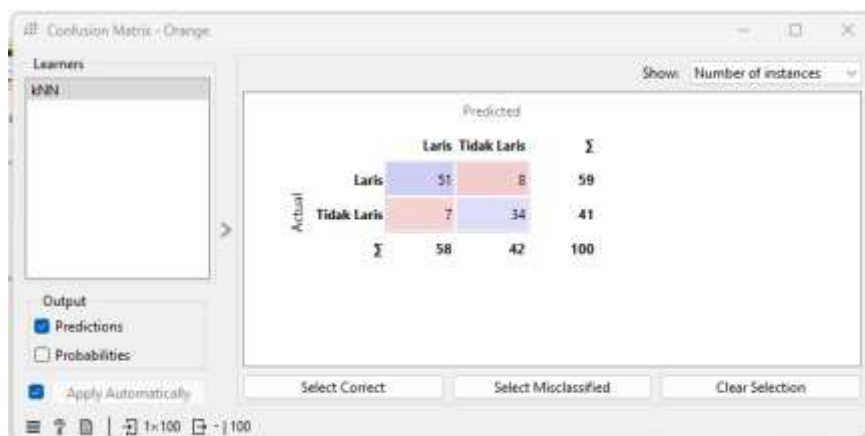
Hasil Evaluasi

Hasil evaluasi pada penelitian ini dilakukan dengan menggunakan widget Test and Score yang mengukur kinerja model berdasarkan metrik akurasi, presisi, recall, dan f1-score. Selain itu, Confusion Matrix digunakan untuk memberikan gambaran lebih jelas mengenai distribusi prediksi yang benar dan salah, serta untuk menganalisis performa model secara lebih mendalam.



Gambar 3. Hasil Test and Score

Pada gambar di atas, diperlihatkan hasil Test and Score yang diperoleh dari penerapan metode K-Nearest Neighbor (KNN) dalam penelitian ini. Berdasarkan hasil evaluasi, AUC (Area Under Curve) yang diperoleh mencapai 0.948, menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antara kelas Laris dan Tidak Laris. Selain itu, nilai Accuracy (CA) sebesar 0.850, F1-score sebesar 0.850, dan Precision sebesar 0.851 menunjukkan bahwa model memiliki performa yang konsisten dalam hal prediksi yang tepat dan seimbang antara true positive dan false positive. Nilai Recall yang sama, yaitu 0.850, menunjukkan bahwa model juga efektif dalam menangkap hampir semua instance yang seharusnya terklasifikasi sebagai Laris. Terakhir, nilai MCC (Matthews Correlation Coefficient) sebesar 0.691 menunjukkan bahwa model memiliki korelasi yang cukup baik antara prediksi dan nilai aktual, yang mengindikasikan performa yang solid dalam klasifikasi data.



Gambar 4. Hasil Confusion Matrix

Berdasarkan confusion matrix yang diberikan, dapat dihitung metrik evaluasi model k-Nearest Neighbors (kNN) dalam bentuk precision, recall, dan F1-score. Untuk kategori "Laris", precision yang diperoleh adalah sekitar 87,9%, artinya dari semua prediksi "Laris", sekitar 87,9% benar-benar termasuk kategori "Laris". Sementara itu, recall untuk "Laris" sebesar 86,4%, menunjukkan bahwa dari seluruh data yang sebenarnya "Laris", 86,4% berhasil diprediksi dengan benar. Untuk kategori "Tidak Laris", precision mencapai sekitar 81,0%, sedangkan recall-nya sebesar 82,9%, yang berarti model cukup baik dalam mendeteksi data yang benar-benar "Tidak Laris". Jika dihitung F1-score, yaitu harmonisasi antara precision



dan recall, model memiliki performa yang seimbang dengan nilai F1-score tinggi pada kedua kelas. Secara keseluruhan, kinerja model dapat dikatakan cukup baik, namun tetap ada ruang untuk peningkatan, terutama dalam mengurangi kesalahan prediksi silang antar kategori.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan metode *K-Nearest Neighbor* (KNN) dalam memprediksi penjualan roti terlaris di Pabrik Roti SAMAWA Rantauprapat menunjukkan kinerja yang cukup baik. Dari proses klasifikasi terhadap 100 data sampel, diperoleh hasil bahwa sebanyak 59 data termasuk dalam kategori “laris” dan 41 data termasuk dalam kategori “tidak laris”. Evaluasi model dilakukan menggunakan *widgit* Test and Score serta Confusion Matrix, dengan hasil evaluasi menunjukkan performa yang cukup tinggi, yaitu nilai AUC sebesar 0,948, nilai akurasi (CA) sebesar 0,850, nilai F1-score sebesar 0,850, presisi sebesar 0,851, recall sebesar 0,850, dan nilai MCC sebesar 0,691. Hasil ini menunjukkan bahwa model KNN mampu memberikan prediksi yang akurat dan dapat diandalkan untuk membantu pengambilan keputusan terkait pengelolaan penjualan roti di masa mendatang.

Daftar Pustaka

- [1] Maruli Tua Silaen, “Klasifikasi Karakteristik Kepribadian Siswa Berdasarkan the Big Five Personality Dengan Menggunakan Metode K- Nearest Neighbor (Knn),” *Jurnal Informatika dan Rekayasa Elektronik*, vol. 6, no. 1, pp. 121–129, 2023, doi: 10.36595/jire.v6i1.860.
- [2] D. Cahyanti, A. Rahmayani, and S. Ainy, “Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara,” vol. 1, no. 2, pp. 39–43, 2020.
- [3] F. Mar’i, N. Husenti, R. R. Brilian, and A. Gumilang, “Pelatihan Pemahaman Implementasi Teori Euclidean Distance dalam Metode K-Nearest Neighbour (K-NN) untuk Komputasi Cerdas bagi Siswa SMA,” *IJCOSIN: Indonesian Journal of Community Service and Innovation*, vol. 3, no. 2, pp. 42–48, 2023, doi: 10.20895/ijcosin.v3i2.1124.
- [4] R. Bahtiar, “Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor,” *Jurnal Informatika MULTI*, vol. 1, no. 3, pp. 203–214, 2023.
- [5] R. Andrianto and F. Irawan, “Implementasi Metode Regresi Linear Berganda Pada Sistem Prediksi Jumlah Tonase Kelapa Sawit di PT . Paluta Inti Sawit,” *Jurnal Pendidikan Tambusai*, vol. 7, no. 1, pp. 2926–2934, 2023.
- [6] M. Aritonang, “Penerapan Algoritma ID3 dalam Prediksi Kebutuhan Pupuk,” vol. 2, no. 4, pp. 247–253, 2021.
- [7] R. Forest and N. Bayes, “Perbandingan Akurasi , Recall , dan Presisi Klasifikasi pada Algoritma,” vol. 5, no. April, pp. 640–651, 2021, doi: 10.30865/mib.v5i2.2937.
- [8] N. Fatma and S. Z. Harahap, “Analysis of Public Interest in Automatic Motorcycles Using KNN and Neural Network Methods,” vol. 8, no. 2, pp. 1178–1187, 2024.
- [9] N. Widya Utami and M. Artana, “Text Mining Dalam Analisis Sentimen Pembelajaran Daring Di Masa Pandemi Covid 19 Menggunakan Algoritma K-Nearest Neighbor,” *Jurnal Informatika Teknologi dan Sains*, vol. 4, no. 2, pp. 140–148, 2022, doi: 10.51401/jinteks.v4i2.2034.
- [10] R. A. Pangestu and S. Noris, “Analisa Data Mining Prediksi Lelang Suku Cadang Dengan Metode K-NearestNeighbor (Studi Kasus PT. Parmud Jaya Perkasa),” *Jurnal Informatika Multi*, vol. 1, no. 4, pp. 285–295, 2023.
- [11] K. N. Di, K. P. P. Pratama, C. Dua, P. Studi, K. Akuntansi, and S. I. Cirebon, “JURNAL DATA SCIENCE & INFORMATIKA (JDSI) Klasifikasi Pemberian Sanksi Pajak Dengan Metode,” vol. 1, no. 2, pp. 41–45, 2021.
- [12] E. Retnoningsih and R. Pramudita, “Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python,” *Bina Insani Ict Journal*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.
- [13] R. Puspita and A. Widodo, “Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 4, p. 646, 2021, doi: 10.32493/informatika.v5i4.7622.
- [14] P. A. Gatto, R. Maulana Awangga, and R. Andarsyah, “Diagnosis Penyakit Demam Berdarah Menggunakan Naïve Bayes,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 3, pp. 1676–1681, 2023, doi: 10.36040/jati.v7i3.6891.
- [15] A. Setiawan, Abd. Rabi, and Y. S. A. Gumilang, “Pengolahan Citra untuk Sortir Buah Stroberi Berdasarkan Kematangan Menggunakan Algoritma K-Nearst Neighbors (KNN),” *Blend Sains Jurnal Teknik*, vol. 2, no. 4, pp. 322–328, 2024, doi: 10.56211/blendsains.v2i4.551.
- [16] A. Mulyanto, E. Susanti, F. Rossi, W. Wajiran, and R. I. Borman, “Penerapan Convolutional Neural Network (CNN) pada Pengenalan Aksara Lampung Berbasis Optical Character Recognition (OCR),” *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 7, no. 1, p. 52, 2021, doi: 10.26418/jp.v7i1.44133.





- [17] M. R. A. Yudianto, K. Kusri, and H. Al Fatta, “Analisis Pengaruh Tingkat Akurasi Klasifikasi Citra Wayang dengan Algoritma Convolutional Neural Network,” *Jurnal Teknologi Informasi*, vol. 4, no. 2, pp. 182–191, 2020, doi: 10.36294/jurti.v4i2.1319.
- [18] A. Ridwan, “Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus,” *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020, doi: 10.47970/siskom-kb.v4i1.169.
- [19] Z. Budiarto, H. Listiyono, and A. Karim, “Optimizing LSTM with Grid Search and Regularization Techniques to Enhance Accuracy in Human Activity Recognition,” *Journal of Applied Data Sciences*, vol. 5, no. 4, pp. 2002–2014, Nov. 2024, doi: 10.47738/jads.v5i4.433.
- [20] B. Bangun and A. K. Karim, “Pengembalian Data Yang Hilang Pada Dataset Dengan Menggunakan Algoritma K-Nearest Neighbor Imputation Data Mining,” *Jurnal Media Informatika Budidarma*, vol. 8, no. 3, p. 1706, 2024, doi: 10.30865/mib.v8i3.8014.
- [21] A. Karim, “Penerapan Algoritma Entropy dan Aras Menentukan Desa Terbaik Di Pemerintah Kabupaten Labuhanbatu,” vol. 3, no. 1, pp. 33–43, 2022.
- [22] A. Karim, “Implementation of the Multi-Objective Optimization Method on the Basic of Ratio Analysis (MOORA) and Entropy Weighting in New Employee Recruitment,” vol. 5, no. 2, pp. 704–712, 2024, doi: 10.47065/josh.v5i2.4859.
- [23] A. Karim, “Sistem Pendukung Keputusan Penerimaan Analisis Di Pusat Penelitian Kelapa Sawit Menggunakan Metode Complex Proportional Assessment (Copras),” *Buletin Ilmiah Informatika Teknologi*, vol. 2, no. 1, pp. 32–42, [Online]. Available: <https://ejurnal.amikstiekomsu.ac.id/index.php/BIIT>
- [24] A. Karim, “Clusterisasi Tingkat Pengangguran Terbuka Menurut Provinsi di Indonesia Menggunakan Algoritma K-Medoids,” 2024, doi: 10.47065/bits.v6i3.6198.
- [25] Abdul Karim, “Implementasi Metode Multi-Objective Optimization On The Basis Of Ratio Analysis dalam Seleksi Mahasiswa Program Indonesia Pintar,” *Bulletin of Computer Science Research*, vol. 3, no. 5, pp. 351–356, 2023, doi: 10.47065/bulletincsr.v3i5.283.

