

PEMODELAN PREDIKSI PENJUALAN PUPUK MENGGUNAKAN NEURAL NETWORK BERDASARKAN DATA HISTORIS PENJUALAN

Mohamad Rizal Rivan Amin, Imam Marzuki*, Hermin Arista

Program Studi Teknik Elektro, Fakultas Teknik dan Informatika, Universitas Panca Marga
Jalan Yos Sudarso No 107 Pabean, Dringu, Probolinggo 67271
imam@upm.ac.id

ABSTRAK

Sektor pertanian di Indonesia memiliki ketergantungan yang tinggi terhadap ketersediaan pupuk sebagai sarana produksi utama. Namun, ketidakakuratan dalam peramalan kebutuhan dan penjualan pupuk seringkali menyebabkan kelebihan atau kekurangan stok yang berdampak signifikan pada efisiensi distribusi dan pendapatan distributor. Permasalahan ini memerlukan pendekatan analitik berbasis data yang mampu mengidentifikasi pola historis penjualan secara akurat. Penelitian ini bertujuan untuk menganalisis dan membangun model prediksi penjualan pupuk menggunakan metode Neural Network yang diimplementasikan pada perangkat lunak Orange Data Mining. Metode penelitian yang digunakan meliputi pengumpulan data historis penjualan pupuk, pra-pemrosesan data, pembangunan model Neural Network dengan arsitektur backpropagation, serta evaluasi menggunakan cross-validation dengan pembagian data latih dan data uji sebesar 80:20. Visualisasi alur analisis dilakukan secara visual menggunakan widget pada antarmuka Orange Data Mining. Hasil pengujian menunjukkan bahwa model Neural Network mampu memprediksi penjualan pupuk dengan akurasi yang tinggi, ditandai oleh nilai Root Mean Square Error (RMSE) yang rendah, sehingga model ini dinilai layak digunakan sebagai alat bantu pengambilan keputusan dalam manajemen persediaan dan strategi distribusi pupuk.

Kata kunci : Data Mining, Neural Network, Orange Data Mining, Prediksi Penjualan, Pupuk, Backpropagation

1. PENDAHULUAN

Sektor pertanian merupakan salah satu tulang punggung perekonomian Indonesia yang menopang ketahanan pangan nasional sekaligus menjadi sumber penghidupan bagi jutaan petani di berbagai daerah. Keberhasilan sektor ini tidak terlepas dari ketersediaan sarana produksi yang memadai, salah satunya adalah pupuk, yang berfungsi menjaga kesuburan tanah dan meningkatkan hasil panen secara signifikan [1]. Pupuk, baik yang bersifat subsidi melalui skema Rencana Definitif Kebutuhan Kelompok (RDKK) maupun non-subsidi, menjadi komoditas strategis yang ketersediaannya harus dikelola dengan cermat oleh distributor agar dapat memenuhi kebutuhan petani secara tepat waktu dan tepat jumlah. Ketersediaan pupuk yang memadai sangat bergantung pada kemampuan distributor dalam membaca pola permintaan pasar dan merencanakan persediaan secara akurat, sehingga diperlukan suatu sistem pendukung keputusan yang mampu membantu proses perencanaan dan pengendalian stok agar distribusi dapat berjalan secara efektif dan efisien [2].

Dalam praktiknya, distributor pupuk masih kerap dihadapkan pada persoalan ketidakpastian jumlah permintaan pada periode mendatang. Pendekatan peramalan konvensional yang hanya mengandalkan intuisi, pengalaman, atau perkiraan subjektif petugas lapangan terbukti rentan menghasilkan estimasi yang kurang akurat. Akibat dari estimasi yang meleset tersebut, distributor dapat terjebak pada dua kondisi yang sama-sama merugikan, yaitu kelebihan stok (overstock) yang membebani biaya penyimpanan dan risiko kerusakan barang, atau sebaliknya kekurangan stok (stockout) yang menghambat distribusi pupuk ke

petani pada saat dibutuhkan dan berpotensi menurunkan tingkat kepuasan pelanggan [3]. Kondisi ini menegaskan bahwa metode peramalan berbasis pengalaman semata sudah tidak lagi memadai untuk menjawab dinamika permintaan pupuk yang dipengaruhi oleh banyak variabel sekaligus, seperti musim tanam, luas lahan, harga jual, dan pola penjualan periode sebelumnya.

Di sisi lain, setiap transaksi penjualan yang dilakukan oleh distributor pupuk pada dasarnya menghasilkan jejak data yang terus terakumulasi dari waktu ke waktu. Data historis tersebut sesungguhnya menyimpan pola dan informasi berharga yang, apabila digali secara tepat, dapat dijadikan dasar pengambilan keputusan yang jauh lebih objektif dibandingkan estimasi manual. Untuk menggali informasi tersembunyi dari tumpukan data tersebut, diperlukan pendekatan data mining, yaitu proses menemukan pola, hubungan, dan pengetahuan baru dari sekumpulan data dalam jumlah besar melalui kombinasi teknik statistik, kecerdasan buatan, dan pembelajaran mesin [4]. Melalui penerapan data mining, data historis penjualan pupuk dapat dianalisis secara sistematis untuk menghasilkan prediksi yang lebih akurat, sehingga distributor dapat menyusun strategi distribusi dan perencanaan persediaan yang lebih terukur.

Salah satu metode pembelajaran mesin yang banyak terbukti efektif untuk persoalan prediksi dan klasifikasi adalah Neural Network atau Jaringan Syaraf Tiruan. Metode ini merupakan bagian dari kecerdasan buatan (Artificial Intelligence) yang meniru cara kerja jaringan saraf biologis manusia dalam memproses informasi, sehingga mampu

mengenali pola-pola kompleks dan hubungan nonlinier yang sulit ditangkap oleh metode statistik konvensional [5], [6]. Karena kemampuannya tersebut, Neural Network telah banyak diadopsi dalam berbagai bidang aplikasi, mulai dari prediksi penjualan, peramalan permintaan, analisis keuangan, hingga sistem pendukung keputusan berbasis data [7]. Pada kasus data penjualan pupuk yang memiliki distribusi kelas tidak seimbang dan dipengaruhi banyak faktor eksternal, kemampuan Neural Network dalam mempelajari pola nonlinier menjadi sangat relevan untuk diuji dan dianalisis lebih lanjut.

Agar implementasi metode Neural Network dapat dilakukan secara efisien tanpa memerlukan keahlian pemrograman yang mendalam, penelitian ini memanfaatkan perangkat lunak Orange Data Mining. Orange merupakan aplikasi analisis data berbasis visual yang bersifat open source, di mana pengguna dapat merangkai alur kerja analisis data melalui serangkaian widget yang saling terhubung, mulai dari proses pemuatan data, pra-pemrosesan, pembangunan model, hingga evaluasi performa [8]. Pendekatan berbasis widget ini memungkinkan proses eksperimen model dilakukan secara lebih cepat, transparan, dan mudah direplikasi, sehingga sangat sesuai untuk kebutuhan penelitian terapan di bidang prediksi penjualan.

Beberapa penelitian terdahulu telah menunjukkan bahwa metode Neural Network mampu menghasilkan tingkat akurasi yang tinggi dalam berbagai persoalan prediksi penjualan dibandingkan metode konvensional [9]. Namun demikian, penerapan metode tersebut secara spesifik pada data penjualan pupuk masih relatif terbatas dan memerlukan kajian lebih lanjut, mengingat karakteristik data penjualan pupuk yang unik dan dipengaruhi oleh berbagai faktor seperti musim tanam, kondisi cuaca, kebutuhan petani, kebijakan distribusi pemerintah, serta ketidakseimbangan jumlah data antar kategori penjualan. Kesenjangan inilah yang mendasari pentingnya dilakukan kajian empiris untuk mengukur sejauh mana kemampuan Neural Network dalam memprediksi kategori penjualan pupuk secara akurat dan andal.

Berdasarkan uraian permasalahan dan kesenjangan penelitian tersebut, penelitian ini dilakukan dengan tujuan untuk menganalisis dan membangun model klasifikasi penjualan pupuk menggunakan metode Neural Network berbasis arsitektur backpropagation yang diimplementasikan pada Orange Data Mining. Model yang dibangun diharapkan mampu mengklasifikasikan data penjualan ke dalam kategori Rendah, Sedang, dan Tinggi dengan tingkat akurasi yang tinggi, sehingga dapat dijadikan sebagai alat bantu pengambilan keputusan yang objektif bagi distributor dalam merencanakan persediaan dan menyusun strategi distribusi pupuk yang lebih efektif dan efisien.

2. TINJAUAN PUSTAKA

Data Mining merupakan proses penggalian informasi, pola, dan pengetahuan yang tersembunyi dari sekumpulan data dalam jumlah besar. Teknik ini memanfaatkan berbagai metode statistik, kecerdasan buatan, dan pembelajaran mesin untuk menghasilkan informasi yang dapat digunakan dalam pengambilan keputusan. Dalam dunia bisnis, data mining sering diterapkan untuk analisis perilaku pelanggan, prediksi penjualan, klasifikasi data, serta identifikasi pola transaksi. Melalui proses data mining, data historis yang sebelumnya hanya menjadi arsip dapat diolah menjadi informasi yang bernilai dan mendukung strategi organisasi.

Tahapan dalam data mining umumnya meliputi proses pengumpulan data, pembersihan data (data cleaning), transformasi data, pemodelan, evaluasi, dan interpretasi hasil. Setiap tahapan memiliki peran penting dalam menghasilkan model yang akurat dan dapat dipercaya. Kualitas data yang digunakan sangat mempengaruhi hasil analisis sehingga diperlukan proses pra-pemrosesan yang baik sebelum dilakukan pembentukan model prediksi. Dalam penelitian ini, data mining digunakan untuk menganalisis data historis penjualan pupuk guna menghasilkan prediksi penjualan pada periode berikutnya.

2.1. Data Mining

Data Mining merupakan proses penggalian informasi, pola, dan pengetahuan yang tersembunyi dari sekumpulan data dalam jumlah besar. Teknik ini memanfaatkan berbagai metode statistik, kecerdasan buatan, dan pembelajaran mesin untuk menghasilkan informasi yang dapat digunakan dalam pengambilan keputusan [4]. Dalam dunia bisnis, data mining sering diterapkan untuk analisis perilaku pelanggan, prediksi penjualan, klasifikasi data, serta identifikasi pola transaksi. Tahapan dalam data mining umumnya meliputi proses pengumpulan data, pembersihan data (data cleaning), transformasi data, pemodelan, evaluasi, dan interpretasi hasil, yang secara keseluruhan dikenal sebagai siklus Knowledge Discovery in Databases (KDD) [9]. Setiap tahapan dalam siklus tersebut memiliki peran penting dan saling memengaruhi: kualitas data pada tahap pembersihan akan menentukan keandalan pola yang ditemukan pada tahap pemodelan, sehingga keduanya tidak dapat dipisahkan dalam proses analisis data yang baik.

2.2. Prediksi Penjualan

Prediksi penjualan merupakan suatu proses memperkirakan jumlah penjualan yang akan terjadi pada masa mendatang berdasarkan data historis dan faktor-faktor yang memengaruhinya. Prediksi yang akurat sangat penting bagi perusahaan karena dapat membantu dalam perencanaan produksi, pengelolaan persediaan, pengaturan distribusi, dan pengambilan keputusan strategis lainnya [3]. Kesalahan dalam prediksi dapat menyebabkan ketidakseimbangan

antara permintaan dan ketersediaan produk sehingga berpotensi menimbulkan kerugian finansial maupun operasional. Dalam konteks distribusi pupuk, prediksi penjualan memiliki peran yang sangat strategis karena kebutuhan pupuk dipengaruhi oleh berbagai faktor yang saling berkaitan, seperti musim tanam, luas lahan pertanian, kondisi cuaca, dan tingkat permintaan petani [1]. Karena keterkaitan faktor-faktor tersebut umumnya bersifat nonlinier, dibutuhkan metode prediksi yang mampu menangkap hubungan kompleks tersebut, salah satunya adalah Neural Network yang diuraikan pada subbab berikut.

2.3. Neural Network dan Algoritma Backpropagation

Neural Network atau Jaringan Syaraf Tiruan (JST) adalah sistem komputasi yang terinspirasi dari struktur jaringan saraf biologis pada otak manusia. JST terdiri atas sejumlah unit pemroses yang disebut neuron yang tersusun dalam lapisan masukan (input layer), satu atau lebih lapisan tersembunyi (hidden layer), dan lapisan keluaran (output layer) [5]. Setiap neuron pada suatu lapisan terhubung dengan neuron pada lapisan berikutnya melalui bobot (weight) yang menentukan kekuatan koneksi antar neuron, dan proses pembelajaran JST pada dasarnya adalah proses pencarian nilai bobot optimal yang meminimalkan kesalahan prediksi. Algoritma yang paling umum digunakan untuk mencari bobot optimal tersebut adalah backpropagation, yang bekerja dengan meminimalkan fungsi error melalui penyesuaian bobot secara iteratif. Proses pembelajaran pada backpropagation terdiri atas dua fase utama, yaitu forward pass untuk menghitung keluaran jaringan berdasarkan bobot saat ini, dan backward pass untuk menyalurkan kembali nilai error guna memperbaiki bobot pada setiap lapisan [6].

2.4. Metrik Evaluasi Klasifikasi

Setelah model Neural Network dilatih, diperlukan tahap evaluasi untuk mengukur sejauh mana model mampu mengklasifikasikan data dengan benar. Evaluasi model klasifikasi umumnya dilakukan menggunakan metrik yang diturunkan dari confusion matrix, yaitu tabel yang menggambarkan perbandingan antara kelas aktual dan kelas hasil prediksi model. Dari confusion matrix tersebut diperoleh empat nilai dasar, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN), yang menjadi komponen pembentuk seluruh metrik evaluasi [11].

2.5. Orange Data Mining

Agar tahapan pemodelan dan evaluasi yang telah diuraikan dapat diimplementasikan secara praktis, penelitian ini memanfaatkan Orange Data Mining sebagai lingkungan kerja utama. Orange Data Mining merupakan perangkat lunak analisis data berbasis visual yang bersifat open source dan dikembangkan oleh Universitas Ljubljana [8]. Perangkat lunak ini

memungkinkan pengguna membangun alur kerja analisis data menggunakan komponen visual yang disebut widget tanpa memerlukan keahlian pemrograman yang mendalam. Berbagai widget tersedia untuk mendukung proses impor data, pra-pemrosesan, pembangunan model, evaluasi, dan visualisasi hasil. Orange mendukung berbagai algoritma pembelajaran mesin termasuk Neural Network, Decision Tree, Random Forest, dan Support Vector Machine (SVM), sehingga memungkinkan peneliti membandingkan beberapa model secara cepat dalam satu lingkungan kerja yang sama [8], [12].

Berdasarkan seluruh kajian teori di atas, dapat disimpulkan bahwa kombinasi antara metode Neural Network dengan algoritma backpropagation, metrik evaluasi confusion matrix yang komprehensif, serta lingkungan implementasi Orange Data Mining yang efisien membentuk satu kerangka kerja yang saling melengkapi untuk menyelesaikan persoalan klasifikasi penjualan pupuk pada penelitian ini. Kerangka kerja tersebut selanjutnya diterjemahkan ke dalam tahapan metodologi penelitian yang dijabarkan pada Bab 3.

3. METODE PENELITIAN

Penelitian ini menerapkan pendekatan data mining untuk melakukan klasifikasi kategori penjualan pupuk menggunakan metode Neural Network berbasis algoritma backpropagation. Data yang digunakan merupakan data historis penjualan pupuk yang terdiri dari beberapa atribut, yaitu Nama Barang, Stok Awal, Penjualan RDKK, Penjualan Non-RDKK, Harga Jual, dan Jumlah Uang Penjualan, dengan target klasifikasi berupa kategori penjualan Rendah, Sedang, dan Tinggi. Seluruh proses pengolahan data, pembangunan model, hingga evaluasi dilakukan menggunakan perangkat lunak Orange Data Mining yang menyediakan lingkungan analisis berbasis visual melalui rangkaian widget yang saling terhubung. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data, pembagian data latih dan data uji, perancangan model Neural Network, implementasi model pada Orange Data Mining, serta evaluasi performa model menggunakan metrik Classification Accuracy (CA), Precision, Recall, dan F1-Score. Rangkaian tahapan tersebut dirancang untuk menghasilkan model klasifikasi yang mampu mengidentifikasi kategori penjualan pupuk secara akurat sehingga dapat digunakan sebagai alat bantu pengambilan keputusan dalam pengelolaan persediaan dan distribusi pupuk.

3.1. Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan data mining yang bertujuan membangun model klasifikasi penjualan pupuk menggunakan metode Neural Network pada perangkat lunak Orange Data Mining. Pendekatan kuantitatif dipilih karena penelitian memanfaatkan data numerik hasil transaksi penjualan sebagai dasar pembentukan model klasifikasi. Model yang dibangun digunakan

untuk mengelompokkan data penjualan ke dalam kategori Rendah, Sedang, dan Tinggi berdasarkan pola yang terbentuk dari data historis.

Metode Neural Network dipilih karena memiliki kemampuan dalam mempelajari hubungan nonlinier antar variabel dan mampu menghasilkan model klasifikasi dengan tingkat akurasi yang tinggi. Seluruh proses analisis dilakukan menggunakan Orange Data Mining yang menyediakan lingkungan pengolahan data secara visual sehingga mempermudah proses implementasi, evaluasi, dan interpretasi hasil.

3.2. Tahapan Penelitian

Penelitian dilaksanakan melalui beberapa tahapan yang dilakukan secara sistematis. Tahap pertama adalah identifikasi masalah dan studi literatur untuk memahami permasalahan pengelolaan stok pupuk dan menentukan metode yang sesuai. Tahap kedua adalah pengumpulan data historis penjualan pupuk. Tahap ketiga adalah pra-pemrosesan data yang meliputi pembersihan, transformasi, dan normalisasi data. Tahap keempat adalah pembagian data menjadi data latih dan data uji. Tahap kelima adalah pembangunan model Neural Network menggunakan data latih. Tahap keenam adalah pengujian model menggunakan data uji dan cross validation. Tahap terakhir adalah evaluasi performa model dan analisis hasil klasifikasi.

3.3. Pengumpulan Data

Data yang digunakan merupakan data sekunder yang diperoleh dari catatan transaksi penjualan pupuk pada distributor pupuk. Dataset terdiri dari 360 data yang mencakup atribut Nama Barang, Stok Awal, Penjualan RDKK, Penjualan Non-RDKK, Harga Jual, Jumlah Uang Penjualan, dan Kategori Penjualan.

Data dikumpulkan dalam format Microsoft Excel (.xlsx) kemudian diimpor ke Orange Data Mining menggunakan widget File. Selanjutnya dilakukan pemeriksaan kualitas data untuk memastikan tidak terdapat data yang hilang, duplikasi data, maupun ketidaksesuaian format data yang dapat memengaruhi proses pelatihan model.

3.4. Pra-pemrosesan Data

Pra-pemrosesan data dilakukan untuk meningkatkan kualitas data sebelum digunakan pada proses klasifikasi. Tahapan ini terdiri dari pembersihan data, transformasi atribut kategorikal menjadi numerik, serta normalisasi data.

Normalisasi dilakukan menggunakan metode Min-Max Normalization untuk menyamakan rentang nilai antar atribut sehingga setiap atribut memiliki kontribusi yang seimbang terhadap proses pembelajaran model.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Keterangan:

(x') = nilai hasil normalisasi

(x) = nilai data asli

(x_{\min}) = nilai minimum atribut

(x_{\max}) = nilai maksimum atribut

Tahap ini dilakukan menggunakan widget Select Columns, Edit Domain, dan Preprocess pada Orange Data Mining.

3.5. Pembagian Data

Setelah proses pra-pemrosesan selesai, data dibagi menjadi dua kelompok yaitu data latih dan data uji. Pembagian dilakukan menggunakan rasio 80:20. Sebanyak 288 data digunakan sebagai data latih untuk membangun model, sedangkan 72 data digunakan sebagai data uji untuk mengukur kemampuan model dalam mengklasifikasikan data yang belum pernah dipelajari sebelumnya.

Pembagian data dilakukan menggunakan widget Data Sampler pada Orange Data Mining.

3.6. Pembangunan Model Neural Network

Model klasifikasi dibangun menggunakan algoritma Neural Network dengan satu hidden layer yang terdiri dari 50 neuron. Fungsi aktivasi yang digunakan adalah Rectified Linear Unit (ReLU), sedangkan proses optimasi menggunakan algoritma Adam. Jumlah maksimum iterasi pelatihan ditetapkan sebanyak 200 iterasi.

Proses perhitungan neuron pada Neural Network dinyatakan sebagai berikut:

$$z = \sum_{i=1}^n w_i x_i + b \quad (2)$$

Keterangan:

(z) = nilai masukan neuron

(w_i) = bobot

(x_i) = atribut input

(b) = bias

Fungsi aktivasi ReLU digunakan untuk menentukan nilai keluaran neuron sebagaimana ditunjukkan pada Persamaan (3).

$$f(x) = \max(0, x) \quad (3)$$

Konfigurasi model Neural Network yang digunakan meliputi 50 neuron pada hidden layer, fungsi aktivasi ReLU, optimizer Adam, nilai regularisasi sebesar 0,0001, dan maksimum iterasi sebanyak 200 epoch.

3.7. Implementasi pada Orange Data Mining

Implementasi model dilakukan menggunakan workflow Orange Data Mining yang terdiri dari widget File, Select Columns, Edit Domain, Preprocess, Data Sampler, Neural Network, Test and Score, Predictions, dan Confusion Matrix.

Alur kerja dimulai dengan pemuatan data menggunakan widget File, kemudian dilanjutkan dengan proses pra-pemrosesan data. Data yang telah diproses dibagi menjadi data latih dan data uji sebelum dimasukkan ke dalam model Neural Network. Selanjutnya model dievaluasi menggunakan widget Test and Score dan hasil klasifikasi divisualisasikan menggunakan widget Predictions dan Confusion Matrix.

3.8. Evaluasi Model

Evaluasi model dilakukan menggunakan metode Cross Validation untuk memperoleh hasil yang lebih objektif dan mengurangi bias akibat pembagian data tertentu. Parameter evaluasi yang digunakan meliputi Classification Accuracy (CA), Precision, Recall, dan F1-Score.

Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

Nilai Accuracy digunakan untuk mengukur tingkat ketepatan klasifikasi secara keseluruhan, Precision digunakan untuk mengukur ketepatan prediksi pada setiap kelas, Recall digunakan untuk mengukur kemampuan model dalam mengenali data yang sebenarnya termasuk dalam suatu kelas, sedangkan F1-Score digunakan untuk menilai keseimbangan antara Precision dan Recall. Semakin tinggi nilai keempat metrik tersebut, maka semakin baik performa model Neural Network yang dihasilkan.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil yang diperoleh dari penerapan metode *Neural Network* untuk mengklasifikasikan data penjualan pupuk ke dalam kategori penjualan menggunakan perangkat lunak *Orange Data Mining*. Hasil yang disajikan meliputi deskripsi data yang digunakan, hasil pengujian model dalam bentuk *confusion matrix*, hasil evaluasi performa model, serta pembahasan terhadap hasil yang diperoleh.

4.1. Deskripsi Data

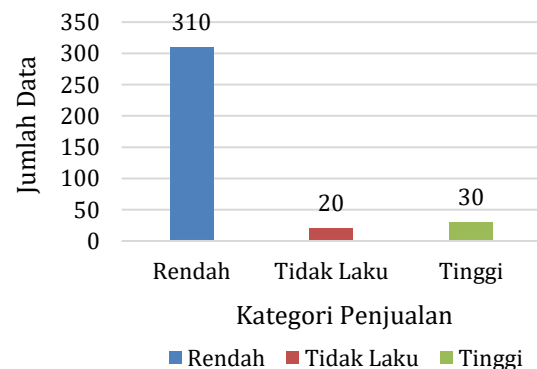
Sebelum dilakukan proses klasifikasi menggunakan Neural Network, terlebih dahulu dilakukan analisis terhadap karakteristik dataset yang digunakan dalam penelitian. Dataset terdiri dari 360 data penjualan pupuk yang diperoleh dari catatan transaksi distributor. Setiap data memiliki atribut Nama Barang, Stok Awal, Penjualan RDKK, Penjualan Non-RDKK, Harga Jual, dan Jumlah Uang Penjualan. Berdasarkan nilai Jumlah Uang Penjualan, data dikelompokkan ke dalam tiga kategori kelas yaitu Rendah, Sedang, dan Tinggi.

Tabel 3 menyajikan distribusi jumlah data pada masing-masing kategori penjualan yang digunakan sebagai label kelas dalam proses klasifikasi.

Tabel 1. Distribusi Kategori Penjualan

Kategori	Jumlah Data	Persentase (%)
Rendah	20	5,6
Sedang	310	86,1
Tinggi	30	8,3
Total	360	100

Berdasarkan Tabel 3, dapat diketahui bahwa distribusi data pada setiap kategori tidak seimbang. Kategori Sedang mendominasi dataset dengan jumlah 310 data atau 86,1% dari keseluruhan data. Sementara itu kategori Rendah hanya terdiri dari 20 data atau 5,6%, sedangkan kategori Tinggi sebanyak 30 data atau 8,3%. Ketidakseimbangan distribusi data tersebut berpotensi memengaruhi kemampuan model dalam mengenali kelas minoritas sehingga perlu diperhatikan pada saat interpretasi hasil klasifikasi.



Gambar 1. Grafik Distribusi Kategori Penjualan Pupuk

Berdasarkan Gambar 1, distribusi data penjualan pupuk menunjukkan bahwa kategori Sedang merupakan kelas yang paling dominan dengan jumlah 310 data atau 86,1% dari keseluruhan dataset. Sementara itu, kategori Tinggi hanya berjumlah 30 data atau 8,3%, sedangkan kategori Rendah merupakan kelas dengan jumlah data paling sedikit yaitu 20 data atau 5,6%.

Perbedaan jumlah data yang cukup signifikan antar kategori menunjukkan bahwa dataset memiliki kondisi class imbalance (ketidakseimbangan kelas). Kondisi ini menyebabkan proses pelatihan model Neural Network lebih banyak mempelajari karakteristik dari kelas mayoritas, yaitu kategori Sedang, dibandingkan kelas minoritas. Akibatnya, model cenderung memberikan prediksi ke arah kelas yang memiliki jumlah data lebih banyak sehingga kemampuan dalam mengenali kategori Rendah menjadi kurang optimal.

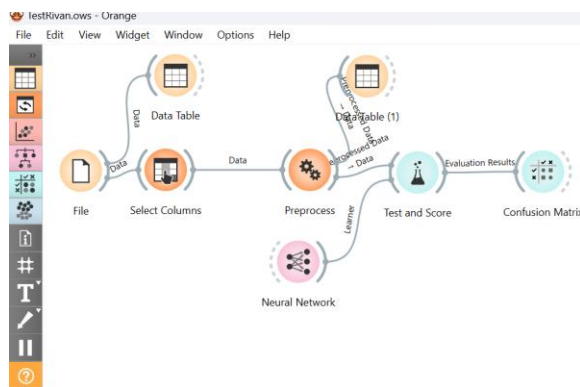
Ketidakseimbangan distribusi data tersebut juga menjadi salah satu penyebab tingginya nilai Classification Accuracy yang diperoleh model. Meskipun akurasi keseluruhan mencapai 91,9%, performa tersebut belum sepenuhnya mencerminkan kemampuan model dalam mengidentifikasi seluruh kategori secara merata. Oleh karena itu, selain menggunakan nilai akurasi, penelitian ini juga mengevaluasi model menggunakan Precision, Recall, F1-Score, serta Confusion Matrix untuk memperoleh gambaran performa klasifikasi yang lebih komprehensif.

Hasil visualisasi ini memperlihatkan bahwa kualitas distribusi data merupakan faktor penting yang

memengaruhi keberhasilan proses pembelajaran Neural Network. Dengan distribusi data yang lebih seimbang, diharapkan kemampuan model dalam mengenali seluruh kategori penjualan dapat meningkat sehingga menghasilkan prediksi yang lebih akurat dan andal untuk mendukung pengambilan keputusan dalam pengelolaan persediaan pupuk.

4.2. Hasil Pengujian Model (Confusion Matrix)

Setelah proses pelatihan model Neural Network selesai dilakukan menggunakan konfigurasi hyperparameter yang telah ditentukan pada Bab 3, tahap berikutnya adalah melakukan pengujian model untuk mengetahui kemampuan klasifikasi terhadap data penjualan pupuk. Pengujian dilakukan menggunakan widget Test and Score serta Confusion Matrix pada Orange Data Mining. Hasil pengujian ditampilkan dalam bentuk confusion matrix yang menunjukkan hubungan antara kelas aktual dan kelas hasil prediksi model.



Gambar 2. Tampilan Workflow Implementasi Model Neural Network pada Orange Data Mining

	Rendah	Sedang	Tinggi	Σ
Actual				
Rendah	0	20	0	20
Sedang	1	307	2	310
Tinggi	0	6	24	30
Σ	1	333	26	360

Gambar 3. Tampilan Widget Confusion Matrix pada Orange Data Mining

Tabel 2. Confusion Matrix Hasil Pengujian Neural Network

Aktual / Prediksi	Rendah	Sedang	Tinggi	Σ
Rendah	0	20	0	20
Sedang	1	307	2	310
Tinggi	0	6	24	30
Σ	1	333	26	360

Berdasarkan Tabel 4, model Neural Network mampu mengklasifikasikan sebagian besar data dengan benar. Pada kategori Sedang, sebanyak 307

dari 310 data berhasil diprediksi secara tepat, sedangkan 2 data diklasifikasikan ke dalam kategori Tinggi dan 1 data ke dalam kategori Rendah. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali karakteristik kategori Sedang.

Pada kategori Sedang, seluruh data aktual diprediksi sebagai kategori Rendah sehingga tidak terdapat satu pun data yang berhasil diklasifikasikan secara benar. Kondisi ini menunjukkan bahwa model masih mengalami kesulitan dalam mengenali pola pada kategori Sedang yang jumlah datanya relatif sedikit dibandingkan kategori lainnya.

Sementara itu, kategori Tinggi menunjukkan performa yang cukup baik dengan 24 data berhasil diprediksi secara benar dari total 30 data aktual. Namun masih terdapat 6 data yang diklasifikasikan sebagai kategori Sedang. Secara keseluruhan, hasil confusion matrix menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali kelas mayoritas, namun masih perlu ditingkatkan untuk mengidentifikasi kelas minoritas secara lebih akurat.

4.3. Hasil Evaluasi Performa Model

Selain menggunakan confusion matrix, performa model Neural Network juga dievaluasi menggunakan beberapa metrik klasifikasi yang terdiri dari Classification Accuracy (CA), Precision, Recall, dan F1-Score. Metrik-metrik tersebut digunakan untuk mengukur tingkat keberhasilan model dalam mengklasifikasikan data penjualan pupuk secara keseluruhan.

Model	CA	F1	Prec	Recall
Neural Network	0.939	0.930	0.933	0.939

Gambar 4. Tampilan Hasil Evaluasi Awal Widget Neural Network dengan Konfigurasi 200 Neuron pada Orange Data Mining

Evaluation results for target (None, show average)				
Model	CA	F1	Prec	Recall
Neural Network	0.919	0.894	0.871	0.919

Gambar 5. Tampilan Konfigurasi dan Hasil Evaluasi Akhir Widget Neural Network (50 Neuron) pada Orange Data Mining

Tabel 3. Hasil Evaluasi Performa Model Neural Network

Model	CA	Precision	Recall	F1-Score
Neural Network	0,919	0,871	0,919	0,894

Berdasarkan Tabel 5, model Neural Network memperoleh nilai Classification Accuracy (CA) sebesar 0,919 atau 91,9%. Nilai tersebut menunjukkan bahwa model mampu melakukan klasifikasi data penjualan pupuk dengan tingkat ketepatan yang sangat baik. Tingginya nilai akurasi menunjukkan bahwa sebagian besar pola pada data berhasil dipelajari selama proses pelatihan model.

Nilai Precision sebesar 0,871 menunjukkan bahwa hasil prediksi yang diberikan model memiliki tingkat ketepatan yang tinggi terhadap kategori yang diprediksi. Selain itu, nilai Recall sebesar 0,919 menunjukkan bahwa model mampu mengenali sebagian besar data aktual yang terdapat pada dataset. Nilai F1-Score sebesar 0,894 mengindikasikan bahwa model memiliki keseimbangan yang baik antara Precision dan Recall.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa metode Neural Network yang diterapkan menggunakan Orange Data Mining mampu memberikan performa klasifikasi yang sangat baik. Dengan tingkat akurasi mencapai 91,9%, model dinilai layak digunakan sebagai alat bantu pengambilan keputusan dalam pengelolaan persediaan dan distribusi pupuk.

4.4. Pembahasan

Secara keseluruhan, hasil pengujian pada Tabel 3 dan Tabel 4 menunjukkan bahwa model Neural Network yang dibangun mampu mengklasifikasikan data penjualan pupuk ke dalam kategori Rendah, Sedang, dan Tinggi dengan tingkat akurasi yang cukup tinggi, yaitu CA sebesar 0,919 atau setara 91,9%. Performa yang baik tersebut sebagian besar didukung oleh tingkat keberhasilan klasifikasi yang tinggi pada kategori Sedang, yang juga merupakan kategori dengan jumlah data terbanyak dalam dataset (310 dari 360 data atau 86,1%). Temuan ini sejalan dengan karakteristik umum algoritma backpropagation, yang cenderung lebih optimal dalam mempelajari pola pada kelas yang memiliki representasi data lebih banyak, sebagaimana juga dilaporkan pada penelitian sebelumnya yang menerapkan Neural Network untuk klasifikasi data pertanian [2] maupun prediksi kebutuhan pupuk bersubsidi [1].

Namun demikian, hasil pada Tabel 3 juga menunjukkan adanya keterbatasan model dalam mengenali data pada kategori Rendah, di mana seluruh 20 data pada kategori tersebut justru diklasifikasikan sebagai kategori Sedang, sehingga Precision dan Recall untuk kategori ini bernilai 0,000 sebagaimana ditunjukkan pada Tabel 5. Kondisi ini diduga kuat disebabkan oleh ketidakseimbangan jumlah data (data

imbalance) antar kategori, di mana jumlah data pada kategori Rendah (20 data, 5,6%) jauh lebih sedikit dibandingkan kategori Sedang (310 data, 86,1%), sehingga selama proses pelatihan model cenderung bias terhadap kategori dengan jumlah data yang lebih dominan. Permasalahan serupa juga banyak dijumpai pada penelitian klasifikasi berbasis Neural Network lainnya, dan secara khusus telah dikaji oleh [7] yang menunjukkan bahwa teknik penyeimbangan data seperti SMOTE mampu memperbaiki kemampuan model dalam mengenali kelas minoritas pada data penjualan ritel.

Apabila ditinjau lebih jauh, kelas Tinggi pada Tabel 3 memperoleh hasil yang relatif lebih baik dibandingkan kelas Rendah, yaitu 24 dari 30 data aktual berhasil diprediksi dengan benar (80%), meskipun 6 data lainnya masih diklasifikasikan sebagai kategori Sedang. Hal ini menunjukkan bahwa batas antara kategori Sedang dan Tinggi relatif lebih mudah dipelajari oleh model dibandingkan batas antara kategori Rendah dan Sedang, kemungkinan karena pola nilai penjualan pada kategori Rendah memiliki kemiripan karakteristik dengan kategori Sedang pada beberapa atribut, seperti Stok Awal dan Penjualan RDKK. Selain itu, nilai Precision (0,871) yang sedikit lebih rendah dibandingkan nilai CA (0,919) dan Recall (0,919) pada Tabel 4 turut mengindikasikan bahwa kesalahan klasifikasi pada kelas dengan jumlah data lebih kecil memberikan kontribusi terhadap penurunan rata-rata Precision secara keseluruhan.

Dari sisi arsitektur model, penggunaan fungsi aktivasi ReLU pada Persamaan (1) terbukti membantu mempercepat proses pembelajaran dan mengurangi risiko vanishing gradient yang umum terjadi pada jaringan dengan banyak lapisan [6]. Demikian pula, penggunaan optimizer Adam yang menggabungkan estimasi momen pertama dan kedua pada Persamaan (3)–(5) berkontribusi pada proses konvergensi yang lebih stabil dibandingkan metode stochastic gradient descent konvensional [10]. Kombinasi kedua komponen arsitektur tersebut menjadi salah satu faktor yang mendukung pencapaian nilai CA dan F1-Score yang tinggi pada penelitian ini, sejalan dengan hasil penelitian terdahulu yang juga menerapkan kombinasi ReLU dan Adam pada klasifikasi data pertanian [2].

Dibandingkan dengan metode klasifikasi konvensional seperti regresi atau metode statistik berbasis aturan, hasil penelitian ini memperkuat temuan [3] yang menyatakan bahwa Neural Network memberikan hasil yang lebih kompetitif pada dataset dengan pola nonlinier dan dipengaruhi banyak variabel, sebagaimana karakteristik data penjualan pupuk yang dipengaruhi faktor musim tanam, harga jual, dan volume penjualan RDKK maupun non-RDKK secara bersamaan. Meskipun demikian, tantangan ketidakseimbangan kelas yang ditemukan pada penelitian ini menunjukkan bahwa tingkat akurasi yang tinggi secara keseluruhan tidak serta-merta mencerminkan kemampuan model yang merata

pada seluruh kategori, sehingga evaluasi per kelas pada Tabel 5 menjadi sama pentingnya dengan evaluasi keseluruhan pada Tabel 4 dalam menilai kelayakan model untuk diterapkan secara praktis.

Berdasarkan keseluruhan pembahasan tersebut, dapat disimpulkan bahwa model Neural Network yang dibangun pada penelitian ini secara umum layak digunakan sebagai alat bantu pengambilan keputusan dalam manajemen persediaan pupuk, khususnya untuk membedakan antara penjualan berkategori Sedang dan Tinggi. Akan tetapi, sebelum diimplementasikan secara penuh pada sistem operasional distributor, diperlukan penanganan lebih lanjut terhadap permasalahan ketidakseimbangan kelas agar kategori Sedang—yang secara bisnis juga penting untuk diidentifikasi guna mencegah penumpukan stok—dapat dikenali oleh model dengan lebih baik.

5. KESIMPULAN DAN SARAN

Berdasarkan seluruh rangkaian tahapan penelitian yang telah dilakukan, mulai dari pengumpulan dan pra-pemrosesan data, perancangan model Neural Network dengan arsitektur backpropagation, hingga pengujian dan evaluasi performa pada Orange Data Mining, dapat ditarik beberapa kesimpulan sebagai berikut.

Pertama, model klasifikasi penjualan pupuk berhasil dibangun menggunakan metode Neural Network dengan konfigurasi 50 neuron pada hidden layer, fungsi aktivasi ReLU (Persamaan 1), dan optimizer Adam (Persamaan 3-5) sebagaimana dirangkum pada Tabel 2. Kedua, hasil evaluasi menggunakan metrik CA, F1-Score, Precision, dan Recall (Persamaan 6-9) menunjukkan performa keseluruhan yang baik, dengan CA = 0,919, F1-Score = 0,894, Precision = 0,871, dan Recall = 0,919 sebagaimana disajikan pada Tabel 4, sehingga secara umum model ini layak dijadikan sebagai alat bantu pengambilan keputusan dalam manajemen persediaan dan strategi distribusi pupuk, khususnya untuk membedakan kategori penjualan Sedang dan Tinggi.

Ketiga, hasil confusion matrix pada Tabel 3 dan evaluasi per kelas pada Tabel 5 mengungkap bahwa performa model tersebut tidak merata pada seluruh kategori. Model menunjukkan kemampuan yang sangat baik dalam mengenali kategori Sedang (F1-Score = 0,955) dan cukup baik pada kategori Tinggi (F1-Score = 0,857), namun gagal sepenuhnya mengenali kategori Rendah (F1-Score = 0,000) akibat ketidakseimbangan jumlah data antar kategori (class imbalance), di mana kategori Rendah hanya direpresentasikan oleh 20 dari 360 data (5,6%). Temuan ini menjadi catatan penting bahwa nilai akurasi keseluruhan yang tinggi tidak selalu mencerminkan kemampuan model yang merata pada seluruh kelas, sehingga evaluasi per kelas tetap diperlukan sebelum model diterapkan secara operasional.

Secara umum, penelitian ini membuktikan bahwa kombinasi metode Neural Network dengan perangkat

lunak Orange Data Mining merupakan pendekatan yang efektif dan efisien untuk membangun model klasifikasi penjualan pupuk, baik dari segi tingkat akurasi yang dicapai maupun dari segi kemudahan implementasi melalui antarmuka visual berbasis widget, sehingga dapat menjadi rujukan bagi distributor maupun peneliti lain yang ingin menerapkan pendekatan serupa pada data penjualan komoditas pertanian lainnya.

Berdasarkan keterbatasan yang ditemukan, khususnya terkait ketidakseimbangan kelas pada kategori Rendah, beberapa saran diajukan bagi penelitian selanjutnya maupun bagi pihak distributor pupuk sebagai berikut.

Pertama, dari sisi metodologis, penelitian selanjutnya disarankan menerapkan teknik penyeimbangan data seperti Synthetic Minority Over-sampling Technique (SMOTE) yang telah terbukti efektif pada kasus serupa [7], atau menerapkan pembobotan kelas (class weighting) pada proses pelatihan model, sehingga kontribusi kategori minoritas seperti Rendah dapat lebih diperhitungkan oleh model selama proses backpropagation berlangsung.

Kedua, perlu dilakukan eksplorasi terhadap metode pembelajaran mesin lain sebagai pembanding, seperti Random Forest, Gradient Boosting, atau Support Vector Machine (SVM) [3], untuk mengetahui apakah metode ensemble atau metode berbasis margin tersebut mampu menangani ketidakseimbangan kelas dengan lebih baik dibandingkan Neural Network pada karakteristik data penjualan pupuk yang sama.

Ketiga, penambahan jumlah dan variasi data, khususnya pada kategori Rendah, perlu dilakukan pada periode pengumpulan data berikutnya agar model memiliki representasi pola yang lebih kaya untuk dipelajari. Selain itu, dapat pula dipertimbangkan penambahan atribut baru yang relevan, seperti data musim tanam, curah hujan, atau wilayah distribusi, untuk memperkaya informasi yang digunakan model dalam proses klasifikasi.

Keempat, bagi pihak distributor pupuk, model Neural Network yang telah dibangun pada penelitian ini dapat mulai dimanfaatkan sebagai alat bantu pendukung keputusan untuk memprediksi kategori penjualan Sedang dan Tinggi dengan tingkat kepercayaan yang tinggi, namun untuk keputusan yang berkaitan dengan produk berisiko Rendah, disarankan agar hasil prediksi model tetap dikombinasikan dengan pertimbangan dan pengalaman lapangan dari petugas distribusi hingga performa model pada kategori tersebut dapat ditingkatkan lebih lanjut.

DAFTAR PUSTAKA

- [1] A. Wahyudi, B. Santoso, dan C. Rahayu, "Penerapan Artificial Neural Network untuk Prediksi Kebutuhan Pupuk Bersubsidi pada Distributor Pertanian," *Jurnal Informatika dan Komputer*, vol. 12, no. 2, hlm. 45–54, 2022.

-
- [2] D. Pratama dan E. Susanti, "Klasifikasi Data Pertanian Menggunakan Backpropagation Neural Network dengan Optimasi Adam," *Jurnal Teknologi Informasi*, vol. 8, no. 1, hlm. 23–31, 2023.
- [3] F. Kurniawan, G. Wibowo, dan H. Nugraha, "Perbandingan Neural Network dan Support Vector Machine dalam Prediksi Penjualan Produk Agribisnis," *Jurnal Ilmu Komputer dan Informasi*, vol. 6, no. 3, hlm. 112–121, 2021.
- [4] I. Fadillah dan J. Permana, "Penerapan Teknik Data Mining dengan Metode Classification untuk Analisis Pola Penjualan Pupuk," *Jurnal Sistem Informasi dan Teknologi*, vol. 5, no. 2, hlm. 78–86, 2022.
- [5] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. New Jersey: Prentice Hall, 2009.
- [6] I. Goodfellow, Y. Bengio, dan A. Courville, *Deep Learning*. Cambridge: MIT Press, 2016.
- [7] M. Prasetyo dan N. Wulandari, "Penanganan Ketidakseimbangan Kelas pada Klasifikasi Data Penjualan Menggunakan SMOTE dan Neural Network," *Jurnal Rekayasa Sistem Informasi*, vol. 9, no. 1, hlm. 34–43, 2024.
- [8] J. Demšar et al., "Orange: Data Mining Toolbox in Python," *Journal of Machine Learning Research*, vol. 14, hlm. 2349–2353, 2013.
- [9] P. N. Tan, M. Steinbach, A. Karpatne, dan V. Kumar, *Introduction to Data Mining*, 2nd ed. New York: Pearson, 2019.
- [10] D. P. Kingma dan J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. International Conference on Learning Representations (ICLR)*, San Diego, 2015. [Online]. Tersedia: <https://arxiv.org/abs/1412.6980>
- [11] M. Sokolova dan G. Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," *Information Processing & Management*, vol. 45, no. 4, hlm. 427–437, 2009.
- [12] Q. Rizaldi dan R. Haryono, "Implementasi Orange Data Mining untuk Klasifikasi Data Transaksi Perdagangan Produk Pertanian," *Jurnal Teknologi dan Manajemen Informatika*, vol. 10, no. 2, hlm. 67–76, 2024.