



Penerapan Data Mining dengan Algoritma C4.5 dan K-nearest Neighbor untuk Prediksi Penjualan Bahan Bangunan Terlaris

Nurhadi Surojudin¹, Muhtajuddin Danny

^{1,2}Universitas Pelita Bangsa

nurhadi@pelitabangsa.ac.id

Abstract

The main problem faced by PT. Surya Kapuas Perkasa is the difficulty in accurately determining the types of building materials with the highest sales levels. Currently, stock determination still relies on manual estimates based on previous sales trends, which are prone to errors and inaccuracies. As a result, the company often faces the risk of overstocking products that are less in demand, or understocking products that are actually in high demand. This condition can impact the sales process, increase storage costs, and reduce customer satisfaction. To overcome this problem, a method is needed that can predict the sales of the best-selling building materials more objectively and based on historical data. This prediction will utilize sales data from the past three years by applying data mining classification techniques using the C4.5 algorithm and K-Nearest Neighbor (K-NN) through the RapidMiner application. With this approach, the company can accurately identify the types of building materials that are most in demand in the market, allowing for more precise and efficient stock management. Based on the research results, four types of building materials were found to be the best-selling out of a total of 16 types analyzed: Light Steel, Brick, Iron, and Cement, with a prediction accuracy rate of 87.16%.

Keywords: Sales Prediction, Building Materials, Data Mining, K-Nearest Neighbor, C4, Rapidminer

Abstrak

Permasalahan utama yang dihadapi PT. Surya Kapuas Perkasa adalah kesulitan dalam menentukan jenis bahan bangunan yang memiliki tingkat penjualan tertinggi secara akurat. Selama ini, penentuan stok barang masih mengandalkan perkiraan manual berdasarkan tren penjualan sebelumnya, yang rentan terhadap kesalahan dan ketidakakuratan. Akibatnya, perusahaan sering dihadapkan pada risiko kelebihan stok (overstock) untuk produk yang kurang diminati, atau kekurangan stok (understock) pada produk yang sebenarnya memiliki permintaan tinggi. Kondisi ini dapat berdampak pada terhambatnya proses penjualan, meningkatnya biaya penyimpanan, dan berkurangnya kepuasan pelanggan. Untuk mengatasi masalah tersebut, diperlukan suatu metode yang mampu melakukan prediksi penjualan bahan bangunan terlaris secara lebih objektif dan berbasis data historis. Prediksi ini akan memanfaatkan data penjualan selama tiga tahun terakhir dengan menerapkan teknik klasifikasi data mining menggunakan algoritma C4.5 dan K-Nearest Neighbor (K-NN) melalui aplikasi RapidMiner. Dengan pendekatan ini, perusahaan dapat mengetahui secara akurat jenis bahan bangunan yang paling diminati pasar, sehingga pengelolaan stok dapat dilakukan secara lebih tepat dan efisien. Berdasarkan hasil penelitian, ditemukan empat jenis bahan bangunan terlaris dari total 16 jenis yang dianalisis, yaitu Baja Ringan, Batu Bata, Besi, dan Semen, dengan tingkat akurasi prediksi sebesar 87,16%.

Kata kunci: Prediksi Penjualan, Bahan Bangunan, Data Mining, K-Nearest Neighbor, C4.5, Rapidminer

INFEB is licensed under a Creative Commons 4.0 International License.



1. Pendahuluan

Industri bahan bangunan merupakan salah satu sektor yang memiliki peranan penting dalam menunjang pertumbuhan infrastruktur dan properti di Indonesia. Permintaan terhadap berbagai jenis bahan bangunan, seperti baja ringan, batu bata, besi, dan semen, cenderung fluktuatif dan dipengaruhi oleh perkembangan proyek konstruksi, tren pasar, serta kondisi ekonomi. Perusahaan penyedia bahan bangunan di wilayahnya, menghadapi tantangan untuk memastikan ketersediaan stok barang sesuai dengan kebutuhan pasar. Selama ini, proses penentuan jumlah dan jenis stok bahan bangunan yang akan disediakan masih mengandalkan perkiraan manual berdasarkan pengalaman dan tren penjualan sebelumnya [1].

Pendekatan konvensional ini memiliki kelemahan, antara lain risiko ketidakakuratan dalam memprediksi

permintaan yang dapat menyebabkan kelebihan stok pada barang yang kurang diminati, atau kekurangan stok pada barang yang memiliki permintaan tinggi [2]. Situasi tersebut tidak hanya menimbulkan kerugian finansial akibat biaya penyimpanan yang membengkak, tetapi juga dapat mengurangi kepuasan pelanggan karena permintaan tidak terpenuhi tepat waktu [3]. Untuk mengatasi permasalahan tersebut, dibutuhkan sebuah metode prediksi penjualan yang akurat dan berbasis data. Pemanfaatan teknik data mining, khususnya metode klasifikasi dengan algoritma K-Nearest Neighbor (K-NN) dan C4.5 [4], dapat menjadi solusi yang efektif. Melalui analisis data penjualan selama tiga tahun terakhir, metode ini diharapkan mampu mengidentifikasi jenis bahan bangunan yang memiliki tingkat penjualan tertinggi secara tepat [5].

Hasil prediksi yang akurat akan membantu perusahaan

dalam mengatur strategi persediaan, mengoptimalkan distribusi, serta meminimalkan risiko kerugian akibat kesalahan estimasi stok [6]. Dalam penerapannya, penelitian prediksi penjualan bahan bangunan terlaris menghadapi beberapa tantangan, seperti kualitas dan kelengkapan data historis penjualan menjadi faktor krusial, mengingat data yang tidak konsisten atau memiliki nilai hilang dapat mempengaruhi hasil prediksi [7]. Dalam dunia bisnis modern, ketersediaan data dalam jumlah besar bukanlah jaminan keberhasilan suatu perusahaan [8]. Data mentah yang tersimpan dalam sistem penjualan hanyalah deretan angka dan catatan transaksi yang belum memberikan makna jika tidak diolah lebih lanjut [9]. Kondisi ini juga berlaku pada PT. Surya Kapuas Perkasa, di mana catatan penjualan bahan bangunan selama tiga tahun terakhir hanya akan menjadi arsip pasif apabila tidak dianalisis secara sistematis. Tanpa proses analisis, perusahaan tidak dapat mengidentifikasi pola penjualan, tren permintaan, maupun jenis produk yang paling diminati pasar.

Dalam era persaingan bisnis yang semakin kompetitif, kemampuan perusahaan untuk mengolah dan memanfaatkan data menjadi salah satu faktor penentu keberhasilan. Data yang tersimpan dalam sistem penjualan tidak hanya menjadi arsip transaksi, tetapi juga mengandung pola dan informasi berharga yang dapat digunakan untuk mendukung pengambilan keputusan strategis. Namun, informasi tersebut sering kali tersembunyi di balik jumlah data yang besar dan kompleks, sehingga sulit diidentifikasi melalui analisis manual.

Data mining hadir sebagai solusi untuk menggali dan mengungkap informasi tersembunyi tersebut melalui penerapan teknik analisis yang sistematis dan berbasis algoritma [10]. Dengan memanfaatkan data mining, perusahaan dapat mengidentifikasi tren, pola pembelian, serta perilaku konsumen yang sebelumnya tidak terlihat. Salah satu metode yang dapat digunakan adalah K-Nearest Neighbor (K-NN), yang bekerja dengan mengklasifikasikan data baru berdasarkan kedekatan jaraknya dengan data historis yang sudah ada [11]. Metode ini sederhana namun efektif, khususnya untuk memprediksi kategori atau klasifikasi produk berdasarkan data penjualan sebelumnya. Masalah spesifik yang dihadapi adalah kesulitan dalam menentukan jenis bahan bangunan yang terlaris, sehingga proses perencanaan stok sering kali tidak optimal.

Perusahaan berisiko mengalami kelebihan stok untuk produk yang kurang diminati atau kekurangan stok pada produk yang memiliki permintaan tinggi. Dengan menerapkan data mining menggunakan metode K-NN pada data penjualan selama tiga tahun terakhir, penelitian ini berupaya memprediksi secara akurat jenis bahan bangunan yang memiliki tingkat penjualan tertinggi [12]. Hasil prediksi ini diharapkan dapat membantu perusahaan dalam mengoptimalkan ketersediaan produk, mengurangi kerugian akibat

kesalahan estimasi stok, serta meningkatkan kepuasan pelanggan [13].

Algoritma C4.5 adalah salah satu metode decision tree (pohon keputusan) yang digunakan untuk melakukan klasifikasi [14]. Prinsip kerjanya adalah membagi data ke dalam cabang-cabang keputusan berdasarkan atribut yang paling informatif, hingga akhirnya menghasilkan kelas atau kategori tertentu. Pada kasus prediksi penjualan bahan bangunan terlaris, C4.5 dapat membantu menemukan atribut penentu (misalnya bulan penjualan, jenis barang, jumlah terjual, atau musim proyek konstruksi) yang paling mempengaruhi tingkat penjualan. Hasilnya adalah decision tree [15] yang dapat digunakan untuk memahami pola penjualan dan membantu proses klasifikasi awal sebelum diterapkan metode lain seperti K-NN.

Algoritma K-NN adalah metode klasifikasi berbasis instance-based learning, di mana data baru diklasifikasikan berdasarkan kesamaan (similarity) dengan data lama. K-NN digunakan sebagai algoritma inti prediksi karena kemampuannya dalam mengklasifikasikan data baru berdasarkan pola penjualan yang sudah ada. Hasil kombinasi ini diharapkan menghasilkan model yang memiliki tingkat akurasi tinggi, terbukti pada penelitian ini dengan capaian akurasi 87,16%, yang berarti model cukup andal untuk mendukung pengambilan keputusan persediaan di perusahaan. Penelitian prediksi penjualan bahan bangunan terlaris pada PT. Surya Kapuas Perkasa menggunakan metode K-Nearest Neighbor (K-NN) memiliki sejumlah kekuatan yang mendukung keberhasilan penerapannya.

Salah satu kekuatan utama adalah kesederhanaan konsep K-NN yang mudah dipahami dan diimplementasikan tanpa memerlukan proses pelatihan model yang kompleks. Metode ini mampu memberikan hasil klasifikasi yang akurat, terutama ketika digunakan pada data penjualan dengan pola yang jelas dan konsisten. Selain itu, fleksibilitas K-NN dalam menangani data dengan berbagai jenis atribut memungkinkan metode ini diterapkan langsung pada dataset historis penjualan selama tiga tahun terakhir tanpa modifikasi yang signifikan. Namun, penerapan K-NN [16] pada dataset yang besar juga menghadirkan tantangan tersendiri. Karena algoritma ini melakukan perhitungan jarak terhadap seluruh data latih setiap kali memproses data baru, waktu komputasi dapat meningkat secara signifikan seiring bertambahnya jumlah data.

Berdasarkan studi literatur, banyak penelitian sebelumnya terkait prediksi penjualan produk telah menggunakan berbagai metode data mining seperti Decision Tree (C4.5), Naïve Bayes, Support Vector Machine (SVM) [17], maupun Artificial Neural Network (ANN) [18]. Namun, sebagian besar penelitian tersebut berfokus pada produk konsumsi umum, seperti makanan, minuman, atau kebutuhan sehari-hari, bukan pada bahan bangunan yang memiliki pola penjualan berbeda karena dipengaruhi oleh musim proyek, tren konstruksi, dan kondisi ekonomi, rentang

waktu data yang pendek, biasanya hanya menggunakan data penjualan satu tahun, sehingga pola musiman dan tren jangka panjang tidak teridentifikasi dengan baik.

Kombinasi algoritma C4.5 dan K-Nearest Neighbor (K-NN), di mana C4.5 digunakan untuk seleksi atribut penting, sedangkan K-NN digunakan sebagai algoritma inti klasifikasi, menghasilkan akurasi yang lebih tinggi (87,16%). Pada RapidMiner [19], modul C4.5 digunakan untuk membangun pohon keputusan, dilanjutkan dengan modul Select Attributes [20] untuk memilih atribut terpilih, dan modul K-NN untuk klasifikasi akhir. Hasil dari proses ini diukur menggunakan metrik evaluasi seperti accuracy, precision, recall, dan confusion matrix untuk mengetahui tingkat keandalan model gabungan yang dibangun. Pendekatan kombinasi C4.5 dan K-NN ini diharapkan dapat menghasilkan model klasifikasi yang tidak hanya akurat tetapi juga efisien, dengan interpretasi yang jelas dari aturan C4.5 sekaligus kekuatan prediksi K-NN [21].

Penelitian ini bertujuan untuk memprediksi penjualan bahan bangunan terlaris pada PT. Surya Kapuas Perkasa berdasarkan data penjualan tiga tahun terakhir. Dengan menerapkan teknik data mining menggunakan kombinasi algoritma C4.5 dan K-Nearest Neighbor (K-NN), penelitian ini berfokus untuk mengidentifikasi jenis produk dengan permintaan tertinggi sehingga perusahaan dapat menentukan strategi pengelolaan stok yang lebih tepat, mengurangi risiko kelebihan atau kekurangan persediaan, mengoptimalkan distribusi produk sesuai permintaan pasar, meningkatkan kepuasan pelanggan melalui ketersediaan produk yang tepat waktu [22].

Sedangkan kontribusi Penelitian ini antara lain: menghadirkan model prediksi berbasis kombinasi algoritma C4.5 dan K-NN yang terbukti efektif dengan akurasi 87,16%, serta memperluas penerapan data mining pada sektor bahan bangunan yang selama ini jarang diteliti dibandingkan sektor ritel umum, memberikan alat bantu pengambilan keputusan bagi PT. Surya Kapuas Perkasa untuk mengelola persediaan secara lebih efisien dan responsif terhadap perubahan permintaan pasar, menangkap pola penjualan jangka panjang melalui penggunaan data historis tiga tahun dan menggabungkannya dengan seleksi atribut menggunakan C4.5 sebelum klasifikasi dengan K-NN, yang menghasilkan prediksi lebih akurat dibandingkan penggunaan algoritma tunggal, dan mendorong implementasi teknologi analisis data pada proses bisnis perusahaan sehingga meningkatkan daya saing di pasar bahan bangunan.

2. Metode Penelitian

Data Selection disajikan pada Tabel 1.

Tabel 1. Data Selection

No	Nama Barang	Satuan	Jumlah	Bulan
1	Alat Listrik	Buah	135	Januari
2	Alat Tukang	Buah	43	Januari
3	Baja Ringan	Batang	646	Januari
4	Batu Bata	Buah	2400	Januari
5	Besi	Batang	406	Januari
6	Cat	Kaleng	330	Januari
7	GRC Board	Lembar	210	Januari
8	Kayu	Kubik	102	Januari
9	Keramik	Buah	168	Januari
10	Kompon	Kg	244	Januari
...
568	Kayu	Kubik	65	Desember
569	Keramik	Buah	36	Desember
570	Kompon	Kg	85	Desember
571	Paku dan Baut	Dus	43	Desember
572	Pasir	Kubik	100	Desember
573	Pipa	Batang	116	Desember
574	Semen	Sak	1700	Desember
575	Thinner	Kaleng	103	Desember
576	Triplek	Lembar	83	Desember

Tabel 1 berisi informasi penjualan berbagai jenis bahan bangunan selama periode satu tahun, dengan rincian setiap bulan. Variabel yang dicatat mencakup: Nama Barang yang merepresentasikan kategori produk, Satuan yang menunjukkan unit pengukuran penjualan, Jumlah yang menyatakan total unit terjual dalam bulan tertentu, serta Bulan yang mengindikasikan waktu transaksi. Tahap *preprocessing* dilakukan untuk pengelompokan jenis penjualan bahan bangunan berdasarkan jumlah penjualan tiap bulan dan tahun untuk memperoleh dalam proses perhitungan prediksi. Setelah data terkelompokan lalu semua dijumlahkan sehingga menjadi data penjualan untuk semua bahan bangunan. Dapat dilihat pada beberapa tabel penjualan bahan bangunan disajikan pada Tabel 2.

Tabel 2. Contoh Data Penjualan Baja Ringan

Bulan	Tahun 2018	Tahun 2019	Tahun 2020
Januari	646	486	452
Februari	598	445	438
Maret	565	517	456
...
Oktober	516	467	425
November	485	458	392
Desember	385	346	325

Berdasarkan Tabel 2. didapatkan hasil pengelompokan penjualan Baja Ringan untuk setiap bulan dan tahun dengan total sebanyak 17070 batang. Selanjutnya contoh data penjualan batu bata disajikan pada Tabel 3.

Tabel 3. Contoh Data Penjualan Batu Bata

Bulan	Tahun 2018	Tahun 2019	Tahun 2020
Januari	2400	2450	2200
Februari	2300	2250	2400
Maret	2500	2200	2500
...
Oktober	2200	1900	2100
November	2150	1850	2400
Desember	1650	1900	1800

Berdasarkan Tabel 3. didapatkan hasil pengelompokan penjualan Batu Bata untuk setiap bulan dan tahun dengan total sebanyak 79300 buah. Selanjutnya contoh data penjualan besi disajikan pada Tabel 4.

Tabel 4. Contoh Data Penjualan Besi

Bulan	Tahun 2018	Tahun 2019	Tahun 2020
Januari	406	406	387
Februari	378	355	365
Maret	380	421	367
...
Oktober	397	412	345
November	423	371	387
Desember	355	297	312

Berdasarkan Tabel 4. didapatkan hasil pengelompokan penjualan Asbes untuk setiap bulan dan tahun dengan total sebanyak 13728 Batang. Selanjutnya contoh data penjualan semen disajikan pada Tabel 5.

Tabel 5. Contoh Data Penjualan Semen

Bulan	Tahun 2018	Tahun 2019	Tahun 2020
Januari	2650	2600	2500
Februari	2400	2250	2350
Maret	2550	2500	2400
...
Oktober	2700	2750	2300
November	2550	2350	2500
Desember	1900	1800	1700

Berdasarkan Tabel 5 didapatkan hasil pengelompokan penjualan Semen untuk setiap bulan dan tahun dengan total sebanyak 85650 sak. Pada tahap *transformation* ini hasil dari pengelompokan data *preprocessing* kemudian digunakan untuk data *training*. Proses pembuatan data *training* berdasarkan data yang ada, data harus di seleksi terlebih dahulu untuk menentukan atribut mana yang dapat mempengaruhi penjualan bahan bangunan terlaris yang disebut data target, dimana data target merupakan data yang berisikan atribut yang akan menjadi atribut yang relevan dan mendukung dalam proses *data mining*. Selanjutnya data training penjualan baja ringan disajikan pada Tabel 6.

Tabel 6. Data Training Penjualan Baja Ringan

No	Bulan -1	Bulan -2	Bulan -3	Bulan -4	Bulan -5	Target
1	646	598	565	612	620	455
2	598	565	612	620	455	575
3	565	612	620	455	575	565
...
17	485	356	523	464	513	467
18	356	523	464	513	467	458
19	523	464	513	467	458	346

Tabel 7. Data Training Penjualan Batu Bata

No	Bulan -1	Bulan -2	Bulan -3	Bulan -4	Bulan -5	Target
1	2400	2300	2500	2300	2550	1600
2	2300	2500	2300	2550	1600	2450
3	2500	2300	2550	1600	2450	2500
...
17	2150	1700	2500	2200	2350	1900
18	1700	2500	2200	2350	1900	1850
19	2500	2200	2350	1900	1850	1900

Tabel 8 Data Training Penjualan Besi

No	Bulan -1	Bulan -2	Bulan -3	Bulan -4	Bulan -5	Target
1	406	378	380	425	412	312
2	378	380	425	412	312	418
3	380	425	412	312	418	398
...
17	435	342	403	395	380	412
18	342	403	395	380	412	371
19	403	395	380	412	371	297

Tabel 9. Data Training Penjualan Semen

No	Bulan -1	Bulan -2	Bulan -3	Bulan -4	Bulan -5	Target
1	2650	2400	2550	2400	2450	1750
2	2400	2550	2400	2450	1750	2700
3	2550	2400	2450	1750	2700	2500
...
17	2450	2000	2550	2550	2400	2750
18	2000	2550	2550	2400	2750	2350
19	2550	2550	2400	2750	2350	1800

Berdasarkan tabel-tabel data *training* yang digunakan pada penelitian ini hanya diambil dari 2 tahun sebelumnya yaitu data penjualan tahun 2018 dan 2019 yang menjadi data *training*. Data *training* dikelompokan menjadi 2 yaitu data masukan dan target. Dimana data masukan merupakan data bulan ke-1 sampai bulan ke-5, sedangkan data target menggunakan data bulan ke-6. Selanjutnya untuk bulan ke-2 hingga bulan ke-6 digunakan sebagai data masukan dengan target atau keluarannya bulan ke-7 dan seterusnya sampai batas data yang ada. Sedangkan untuk data *testing* (uji) yang digunakan untuk prediksi adalah pada Tabel 10.

Tabel 10. Data Testing Penjualan Baja Ringan

Bulan-1	Bulan-2	Bulan-3	Bulan-4	Bulan-5
513	467	458	346	452
467	458	346	452	438
458	346	452	438	456
...
438	425	432	395	425
425	432	395	425	392
432	395	425	392	325

Tabel 11. Data Testing Penjualan Batu Bata

Bulan-1	Bulan-2	Bulan-3	Bulan-4	Bulan-5
2350	1900	1850	1900	2200
1900	1850	1900	2200	2400
1850	1900	2200	2400	2500
...
2350	2200	2300	2300	2100
2200	2300	2300	2100	2400
2300	2300	2100	2400	1800

Tabel 12. Data Testing Penjualan Besi

Bulan-1	Bulan-2	Bulan-3	Bulan-4	Bulan-5
380	412	371	297	387
412	371	297	387	365
371	297	387	365	367
...
371	387	374	411	345
387	374	411	345	387
374	411	345	387	312

Berdasarkan tabel-tabel menunjukkan bahwa data yang digunakan untuk di *test* (uji) adalah data penjualan

tahun 2020 dari bulan januari sampai dengan bulan desember yang akan menghasilkan prediksi penjualan pada tahun 2021.

3. Hasil dan Pembahasan

RapidMiner merupakan salah satu *software data mining* pengolahan data set untuk mencari pola data sesuai dengan tujuan dari pengolahan data tersebut, tidak semua algoritma yang ada dapat sesuai atau dapat mengolah data set yang ada, harus dilakukan penyesuaian pola data dan sesuai dengan tujuan dari pengolahan data tersebut. Langkah awal dalam mengelola data adalah memasukkan data yang ingin diolah kedalam *rapidminer*. Cari operator *read excel* dan masukkan datanya, data yang akan dimasukkan yaitu data *selection*. Pastikan *sheet* data yang ingin dimasukkan benar ditampilkan Gambar 1.

	A	B	C	D
1	No	Name	Category	Price
2	1	Hand Lampu	Light	720
3	2	Hand Lampu	Light	420
4	3	Hand Lampu	Light	320
5	4	Hand Lampu	Light	2000
6	5	Ball	Ball	400
7	6	Ball	Ball	300
8	7	Ball	Ball	220
9	8	Ball	Ball	100
10	9	Ball	Ball	100
11	10	Wingman	Ball	200
12	11	Wingman	Ball	100
13	12	Pen	Pen	90
14	13	Pen	Pen	100

Gambar 1. Pemilihan Data

Pada atribut nama barang pilih sebagai label, lalu tekan *finish* ditampilkan pada gambar 2.

Index	Name	Label	Value	Value
1	Hand Lampu	Label	720	200000
2	Hand Lampu	Label	420	200000
3	Hand Lampu	Label	320	200000
4	Ball	Label	400	200000
5	Ball	Label	300	200000
6	Ball	Label	220	200000
7	Ball	Label	100	200000
8	Ball	Label	100	200000
9	Wingman	Label	200	200000
10	Wingman	Label	100	200000
11	Pen	Label	90	200000
12	Pen	Label	100	200000

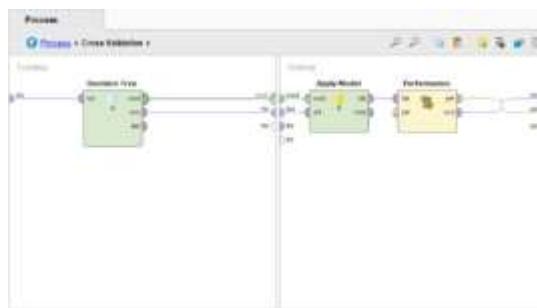
Gambar 2. Pemilihan Atribut Sebagai Label

Cari operator *cross validation* pada kolom pencarian operator, selanjutnya hubungkan kabel-kabelnya dengan benar lalu *double click* pada *cross validation* ditampilkan pada Gambar 3.



Gambar 3. Cross Validation

Dalam proses ini dimasukan operator *Decision Tree*, *Apply Model* dan *Performance* pada *Main Proses*, lalu hubungkan kabel-kabel tersebut secara benar dan tekan tombol *run* atau *play* (segitiga berwarna biru) di tampilkan pada Gambar 4.



Gambar 4. Main Proses Validation

Adapun hasil *accuracy* dari *performance vector* sebesar 87,16% untuk proses yang dilaksanakan pada model diatas dapat ditampilkan pada Gambar 5.

Model	Accuracy	Recall	Precision	F1 Score	Support
proto_0	0.8716	0.8716	0.8716	0.8716	1000
proto_1	0.8716	0.8716	0.8716	0.8716	1000
proto_2	0.8716	0.8716	0.8716	0.8716	1000
proto_3	0.8716	0.8716	0.8716	0.8716	1000
proto_4	0.8716	0.8716	0.8716	0.8716	1000
proto_5	0.8716	0.8716	0.8716	0.8716	1000
proto_6	0.8716	0.8716	0.8716	0.8716	1000
proto_7	0.8716	0.8716	0.8716	0.8716	1000
proto_8	0.8716	0.8716	0.8716	0.8716	1000
proto_9	0.8716	0.8716	0.8716	0.8716	1000
proto_10	0.8716	0.8716	0.8716	0.8716	1000
proto_11	0.8716	0.8716	0.8716	0.8716	1000
proto_12	0.8716	0.8716	0.8716	0.8716	1000
proto_13	0.8716	0.8716	0.8716	0.8716	1000
proto_14	0.8716	0.8716	0.8716	0.8716	1000
proto_15	0.8716	0.8716	0.8716	0.8716	1000
proto_16	0.8716	0.8716	0.8716	0.8716	1000
proto_17	0.8716	0.8716	0.8716	0.8716	1000
proto_18	0.8716	0.8716	0.8716	0.8716	1000
proto_19	0.8716	0.8716	0.8716	0.8716	1000
proto_20	0.8716	0.8716	0.8716	0.8716	1000
proto_21	0.8716	0.8716	0.8716	0.8716	1000
proto_22	0.8716	0.8716	0.8716	0.8716	1000
proto_23	0.8716	0.8716	0.8716	0.8716	1000
proto_24	0.8716	0.8716	0.8716	0.8716	1000
proto_25	0.8716	0.8716	0.8716	0.8716	1000
proto_26	0.8716	0.8716	0.8716	0.8716	1000
proto_27	0.8716	0.8716	0.8716	0.8716	1000
proto_28	0.8716	0.8716	0.8716	0.8716	1000
proto_29	0.8716	0.8716	0.8716	0.8716	1000
proto_30	0.8716	0.8716	0.8716	0.8716	1000
proto_31	0.8716	0.8716	0.8716	0.8716	1000
proto_32	0.8716	0.8716	0.8716	0.8716	1000
proto_33	0.8716	0.8716	0.8716	0.8716	1000
proto_34	0.8716	0.8716	0.8716	0.8716	1000
proto_35	0.8716	0.8716	0.8716	0.8716	1000
proto_36	0.8716	0.8716	0.8716	0.8716	1000
proto_37	0.8716	0.8716	0.8716	0.8716	1000
proto_38	0.8716	0.8716	0.8716	0.8716	1000
proto_39	0.8716	0.8716	0.8716	0.8716	1000
proto_40	0.8716	0.8716	0.8716	0.8716	1000
proto_41	0.8716	0.8716	0.8716	0.8716	1000
proto_42	0.8716	0.8716	0.8716	0.8716	1000
proto_43	0.8716	0.8716	0.8716	0.8716	1000
proto_44	0.8716	0.8716	0.8716	0.8716	1000
proto_45	0.8716	0.8716	0.8716	0.8716	1000
proto_46	0.8716	0.8716	0.8716	0.8716	1000
proto_47	0.8716	0.8716	0.8716	0.8716	1000
proto_48	0.8716	0.8716	0.8716	0.8716	1000
proto_49	0.8716	0.8716	0.8716	0.8716	1000
proto_50	0.8716	0.8716	0.8716	0.8716	1000
proto_51	0.8716	0.8716	0.8716	0.8716	1000
proto_52	0.8716	0.8716	0.8716	0.8716	1000
proto_53	0.8716	0.8716	0.8716	0.8716	1000
proto_54	0.8716	0.8716	0.8716	0.8716	1000
proto_55	0.8716	0.8716	0.8716	0.8716	1000
proto_56	0.8716	0.8716	0.8716	0.8716	1000
proto_57	0.8716	0.8716	0.8716	0.8716	1000
proto_58	0.8716	0.8716	0.8716	0.8716	1000
proto_59	0.8716	0.8716	0.8716	0.8716	1000
proto_60	0.8716	0.8716	0.8716	0.8716	1000
proto_61	0.8716	0.8716	0.8716	0.8716	1000
proto_62	0.8716	0.8716	0.8716	0.8716	1000
proto_63	0.8716	0.8716	0.8716	0.8716	1000
proto_64	0.8716	0.8716	0.8716	0.8716	1000
proto_65	0.8716	0.8716	0.8716	0.8716	1000
proto_66	0.8716	0.8716	0.8716	0.8716	1000
proto_67	0.8716	0.8716	0.8716	0.8716	1000
proto_68	0.8716	0.8716	0.8716	0.8716	1000
proto_69	0.8716	0.8716	0.8716	0.8716	1000
proto_70	0.8716	0.8716	0.8716	0.8716	1000
proto_71	0.8716	0.8716	0.8716	0.8716	1000
proto_72	0.8716	0.8716	0.8716	0.8716	1000
proto_73	0.8716	0.8716	0.8716	0.8716	1000
proto_74	0.8716	0.8716	0.8716	0.8716	1000
proto_75	0.8716	0.8716	0.8716	0.8716	1000
proto_76	0.8716	0.8716	0.8716	0.8716	1000
proto_77	0.8716	0.8716	0.8716	0.8716	1000
proto_78	0.8716	0.8716	0.8716	0.8716	1000
proto_79	0.8716	0.8716	0.8716	0.8716	1000
proto_80	0.8716	0.8716	0.8716	0.8716	1000
proto_81	0.8716	0.8716	0.8716	0.8716	1000
proto_82	0.8716	0.8716	0.8716	0.8716	1000
proto_83	0.8716	0.8716	0.8716	0.8716	1000
proto_84	0.8716	0.8716	0.8716	0.8716	1000
proto_85	0.8716	0.8716	0.8716	0.8716	1000
proto_86	0.8716	0.8716	0.8716	0.8716	1000
proto_87	0.8716	0.8716	0.8716	0.8716	1000
proto_88	0.8716	0.8716	0.8716	0.8716	1000
proto_89	0.8716	0.8716	0.8716	0.8716	1000
proto_90	0.8716	0.8716	0.8716	0.8716	1000
proto_91	0.8716	0.8716	0.8716	0.8716	1000
proto_92	0.8716	0.8716	0.8716	0.8716	1000
proto_93	0.8716	0.8716	0.8716	0.8716	1000
proto_94	0.8716	0.8716	0.8716	0.8716	1000
proto_95	0.8716	0.8716	0.8716	0.8716	1000
proto_96	0.8716	0.8716	0.8716	0.8716	1000
proto_97	0.8716	0.8716	0.8716	0.8716	1000
proto_98	0.8716	0.8716	0.8716	0.8716	1000
proto_99	0.8716	0.8716	0.8716	0.8716	1000
proto_100	0.8716	0.8716	0.8716	0.8716	1000

Gambar 5. Hasil Perhitungan Nilai Akurasi



Gambar 6. Model Pengelompokan Decision Tree

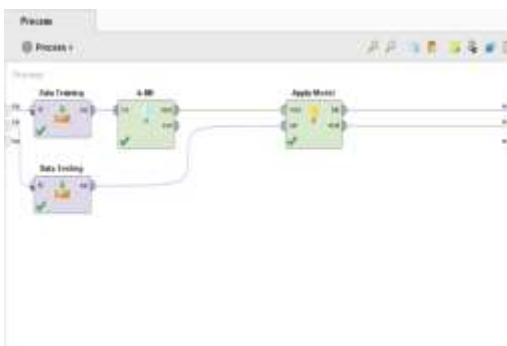
Berdasarkan gambar adalah model pengelompokan nama barang berdasarkan pengklasifikasian pada metoden *Decision Tree* menggunakan *tolls rapidminer*. Setelah melakukan klasifikasi, tahap selanjutnya yang dilakukan adalah cari operator *read excel* pada kolom

pencarian operator kemudian *drag and drop*. Kemudian lakukan *importing* tabel *microsoft excel* kedalam proses untuk membacar data *training* dalam format *excel*, kemudian ganti atribut target menjadi label ditampilkan pada Gambar 7.

The screenshot shows the 'Format your columns' dialog in RapidMiner. It displays a table with 12 columns labeled 'Bulan-1' through 'Bulan-5' and 'Target'. The 'Target' column is highlighted in green. Below the table, there are buttons for 'Import' and 'Cancel'.

Gambar 7. Importing Data Training

Selanjutnya *drag and drop* operator k-NN dengan memasukan nilai $k=3$. Kemudian hubungkan operator *read excel* (data *training*) dengan operator k-NN. Setelah itu hubungkan k-NN dengan operator *apply model*, kemudian ulangi lagi *drag* operator *read excel* (data *testing*) dan *import* tabel data *testing*, sambungkan pada operator *apply model* setelah itu hubungkan pada *result* ditampilkan pada Gambar 8.



Gambar 8. Susunan Operator Algoritma K-NN

Tahap ini adalah tahap terakhir dari proses *data mining* menggunakan *rapidminer* dimana setelah semua operator terhubung kemudian klik tombol *run / play*, maka akan muncul sebuah tab *result*, yang isinya sebuah prediksi dari seluruh data yang memenuhi nilai $k=3$, didapatkan hasil prediksi penjualan ditampilkan pada Gambar 9.

The screenshot shows the 'Result History' table in RapidMiner. It lists 16 rows of data with columns: 'Bulan-1', 'prediksiPenjualan', 'Bulan-2', 'Bulan-3', 'Bulan-4', and 'Bulan-5'. The last row, which corresponds to July, has a value of 534 in the 'prediksiPenjualan' column.

Gambar 9. Hasil Prediksi Penjualan Baja Ringan

Gambar 9 merupakan hasil prediksi penjualan jenis Baja Ringan dimana prediksi penjualan terbanyak pada bulan Juli dengan nilai prediksi sebanyak 534.

The screenshot shows the 'Result History' table in RapidMiner. It lists 12 rows of data with columns: 'Bulan-1', 'prediksiPenjualan', 'Bulan-2', 'Bulan-3', 'Bulan-4', and 'Bulan-5'. The last row, which corresponds to June, has a value of 2318 in the 'prediksiPenjualan' column.

Gambar 10. Hasil Prediksi Penjualan Batu Bata

Gambar 10 merupakan hasil prediksi penjualan jenis Batu Bata dimana prediksi penjualan terbanyak pada bulan Juni dengan nilai prediksi sebanyak 2318.

The screenshot shows the 'Result History' table in RapidMiner. It lists 12 rows of data with columns: 'Bulan-1', 'prediksiPenjualan', 'Bulan-2', 'Bulan-3', 'Bulan-4', and 'Bulan-5'. The last row, which corresponds to March, has a value of 410 in the 'prediksiPenjualan' column.

Gambar 11. Hasil Prediksi Penjualan Besi

Gambar 11 merupakan hasil prediksi penjualan jenis Besi dimana prediksi penjualan terbanyak pada bulan Maret dengan nilai prediksi sebanyak 410.

The screenshot shows the 'Result History' table in RapidMiner. It lists 16 rows of data with columns: 'Bulan-1', 'prediksiPenjualan', 'Bulan-2', 'Bulan-3', 'Bulan-4', and 'Bulan-5'. The last row, which corresponds to March, has a value of 2648 in the 'prediksiPenjualan' column.

Gambar 12. Hasil Prediksi Penjualan Semen

Gambar 12 merupakan hasil prediksi penjualan jenis Semen dimana prediksi penjualan terbanyak pada bulan Maret dengan nilai prediksi sebanyak 2648.

4. Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa penerapan teknik klasifikasi data mining dengan mengombinasikan algoritma C4.5 dan K-Nearest Neighbor (K-NN) melalui aplikasi RapidMiner mampu menjadi solusi efektif bagi PT. Surya Kapuas Perkasa dalam mengatasi permasalahan pengelolaan stok bahan bangunan. Metode ini memungkinkan prediksi jenis bahan bangunan terlaris dilakukan secara objektif dan akurat dengan memanfaatkan data penjualan historis, sehingga meminimalkan risiko terjadinya overstock maupun understock. Dari total 16 jenis bahan

bangunan yang dianalisis, diperoleh empat jenis dengan tingkat permintaan tertinggi, yaitu Baja Ringan, Batu Bata, Besi, dan Semen, dengan akurasi prediksi sebesar 87,16%. Hasil ini membuktikan bahwa integrasi C4.5 sebagai metode seleksi fitur dan K-NN sebagai algoritma klasifikasi mampu meningkatkan efektivitas pengambilan keputusan, mengoptimalkan pengelolaan persediaan, menekan biaya penyimpanan, serta meningkatkan kepuasan pelanggan. Implementasi kombinasi algoritma C4.5 dan K-Nearest Neighbor (K-NN) dalam penelitian ini dilakukan menggunakan workflow di RapidMiner yang tersusun secara sistematis agar menghasilkan proses klasifikasi yang optimal. Tahap awal dimulai dengan import data, di mana dataset dimasukkan ke RapidMiner melalui operator Read Excel atau Read CSV, kemudian dilakukan pengaturan peran atribut (Set Role) untuk menentukan label (atribut target) dan atribut prediktor. Selanjutnya, pada tahap data preprocessing, dilakukan pembersihan data dengan operator Replace Missing Values untuk mengisi nilai yang hilang, normalisasi data menggunakan Normalize atau konversi numerik ke biner melalui Numerical to Binominal jika diperlukan, serta penyaringan data yang tidak relevan menggunakan Filter Examples. Setelah data siap, dilakukan penerapan algoritma C4.5 dengan operator Decision Tree yang menggunakan kriteria Gain Ratio sesuai karakteristik C4.5. Parameter seperti kedalaman pohon, minimal gain, dan confidence level diatur agar menghindari overfitting, sementara hasil model tree dimanfaatkan untuk mengidentifikasi atribut yang paling berpengaruh terhadap klasifikasi. Berdasarkan hasil penelitian tersebut, terdapat beberapa saran yang dapat diberikan untuk pengembangan dan penerapan metode prediksi penjualan bahan bangunan di masa mendatang. Pertama, perusahaan sebaiknya mengimplementasikan sistem prediksi berbasis kombinasi algoritma C4.5 dan K-NN secara berkelanjutan agar pengambilan keputusan pengelolaan stok selalu didasarkan pada data terkini dan akurat. Kedua, penelitian lanjut dapat memperluas cakupan variabel yang digunakan, misalnya dengan memasukkan faktor eksternal seperti tren pasar, kondisi ekonomi, musim proyek konstruksi, dan harga bahan bangunan, sehingga model prediksi dapat lebih adaptif terhadap dinamika permintaan. Ketiga, optimasi parameter pada algoritma C4.5 dan K-NN perlu dilakukan secara berkala melalui parameter tuning untuk menjaga performa model tetap optimal. Keempat, sistem prediksi dapat diintegrasikan dengan perangkat lunak manajemen persediaan (inventory management system) guna mempercepat proses pemesanan ulang, meminimalkan risiko kesalahan, dan meningkatkan efisiensi operasional. Terakhir, penelitian berikutnya dapat membandingkan kinerja kombinasi C4.5-K-NN dengan algoritma lain seperti Random Forest, Gradient Boosting, atau Support Vector Machine untuk menemukan metode yang paling unggul dalam konteks prediksi penjualan bahan bangunan.

Daftar Rujukan

- [1] Everaldo, D., Achmadi, S., & Pranoto, Y. A. (2021). Sistem Informasi Kebutuhan Bahan Pembangunan Rumah Berbasis Website (Studi Kasus : PT. Taniya Multi Properti). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(2), 720–727. DOI: <https://doi.org/10.36040/jati.v5i2.3728>.
- [2] Banjarnahor, J., Reinaldo, E., & Indra, E. (2020). Penerapan Data Mining dengan Algoritma ID3 untuk Memprediksi Penjualan (Studi Kasus : PT. Tata Warna Cipta Perkasa). *Jurnal Sistem Informasi dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, 3(2), 1–6. DOI: <https://doi.org/10.34012/jusikom.v3i2.836> .
- [3] Anggrawan, A., Hairani, H., & Azmi, N. (2022). Prediksi Penjualan Produk Unilever Menggunakan Metode Regresi Linear. *Jurnal Bumigora Information Technology (BITe)*, 4(2), 123–132. DOI: <https://doi.org/10.30812/bite.v4i2.2416> .
- [4] Firmansyah, M. A., Panji Sasmito, A., & Zulfia Zahro', H. (2021). Aplikasi Forecasting Penjualan Bahan Bangunan Menggunakan Metode Trend Moment (Studi Kasus di UD. Hasil Bumi). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(2), 526–533. DOI: <https://doi.org/10.36040/jati.v5i2.3759> .
- [5] Prastiwi, H., Jeny Pricilia, & Errissya Rasywir. (2022). Implementasi Data Mining untuk Menentukan Persediaan Stok Barang di Mini Market Menggunakan Metode K-Means Clustering. *Jurnal Informatika dan Rekayasa Komputer(JAKAKOM)*, 2(1), 141–148. DOI: <https://doi.org/10.33998/jakakom.2022.2.1.34> .
- [6] Fathurrozi, A., Masya, F., & Sugiyatno. (2023). Implementasi Algoritma Apriori untuk Prediksi Transaksi Penjualan Produk pada Aplikasi Point of Sales. *Technomedia Journal*, 8(2), 70–81. DOI: <https://doi.org/10.33050/tmj.v8i2.2004> .
- [7] Gultom, M. M., & Maryam. (2020). Sistem Informasi Penjualan Material Bangunan pada Toko Bangunan Berkah. *Jurnal Teknik Informatika (Jutif)*, 1(2), 79–86. DOI: <https://doi.org/10.20884/1.jutif.2020.1.2.19> .
- [8] Zulkifli, Asmawati.S, & Arnita Irianti. (2022). Penerapan Algoritma Naive Bayes dalam Memprediksi Persediaan Bahan Mebel (Studi Kasus Mebel Usaha Bersama Palipi Soreang). *Journal of Computer and Information System (JCIS)*, 5(1), 57–64. DOI: <https://doi.org/10.31605/jcis.v5i1.1360> .
- [9] Firmansyah, F., & Nurdianwan, O. (2023). Penerapan Data Mining Menggunakan Algoritma Frequent Pattern - Growth untuk Menentukan Pola Pembelian Produk Chemicals. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 547–551. DOI: <https://doi.org/10.36040/jati.v7i1.6371> .
- [10] Soleh, P., Tholib, A., & Hidayat, M. N. F. (2022). Penerapan Data Mining untuk Analisa Pola Pembelian Produk Menggunakan Algoritma Frequent Pattern – Growth. *Rekayasa*, 14(3), 456–460. DOI: <https://doi.org/10.21107/rekayasa.v14i3.11365> .
- [11] Gustipartsani, K., Rahaningsih, N., Danar Dana, R., & Yulia Mustafa, I. (2024). Data Mining Clustering Menggunakan Algoritma K-Means pada Data Kunjungan Wisatawan di Kabupaten Karawang. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3595–3601. DOI: <https://doi.org/10.36040/jati.v7i6.8282> .
- [12] Setianingrum, A., Hindayanti, A., Cahya, D. M., & Purnia, D. S. (2021). Perbandingan Metode Algoritma K-NN & Metode Algoritma C45 pada Analisa Kredit Macet (Studi Kasus PT Tungmung Textil Bintan). *EVOLUSI: Jurnal Sains Dan Manajemen*, 9(2). DOI: <https://doi.org/10.31294/evolusi.v9i2.11166> .
- [13] Ridla, M. A., Bajuri, A., & Ahmad, U. (2023). Implementasi Data Mining terhadap Pola Penjualan Bahan Material Bangunan di TB. Murah Rejeki Menggunakan Algoritma Apriori. *Jurnal SIMADA (Sistem Informasi Dan Manajemen Basis Data)*, 6(2), 92–103. DOI: <https://doi.org/10.30873/simada.v6i2.3800> .
- [14] Tukino, T. (2019). Penerapan Algoritma C4.5 Untuk Memprediksi Keuntungan pada PT SMÖE Indonesia. *JURNAL*

- SISTEM INFORMASI BISNIS, 9(1), 39. DOI: <https://doi.org/10.21456/vol9iss1pp39-46> .
- [15]Arifin, N. B. A. B., & Asmianto, A. (2023). Sistem Prediksi Penjualan Menggunakan Kombinasi Metode Monte Carlo dan Decision Tree Berbasis Website. *MATHunesa: Jurnal Ilmiah Matematika*, 11(2), 274–286. DOI: <https://doi.org/10.26740/mathunesa.v11n2.p274-286> .
- [16]Mujilahwati, S., & Windasari, L. D. (2024). Implementasi Metode k-Nearest Neighbor (k-NN) untuk Memprediksi Penjualan Buah di Indonesia berbasis Website. *Seminar Nasional Teknologi & Sains*, 3(1), 7–14. DOI: <https://doi.org/10.29407/stains.v3i1.4077> .
- [17]Oktaviana Isbirotin, Wiwit Herulambang, Rahmawati Febriyaning Tias, Rangsang, & Ahmad. (2023). Prediction of Skincare Sales Turnover Using the Support Vector Method at the Widya Msglow Sidoarjo Company. *JEECS (Journal of Electrical Engineering and Computer Sciences)*, 8(2), 181–190. DOI: <https://doi.org/10.54732/jeeecs.v8i2.10> .
- [18]Rahmah, S., Witanti, W., & Sabrina, P. N. (2023). Prediksi Penjualan Obat Menggunakan Metode Adaptive Neuro-Fuzzy Inference System (ANFIS). *J I M P - Jurnal Informatika* Merdeka Pasuruan, 7(3), 109. DOI: <https://doi.org/10.51213/jimp.v7i3.733> .
- [19]Nasrullah, A. H. (2018). Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out. *ILKOM Jurnal Ilmiah*, 10(2), 244–250. DOI: <https://doi.org/10.33096/ilkom.v10i2.300.244-250> .
- [20]Wariyanti Nugroho Putri, Made Hanindia Prami Swari, & Retno Mumpuni. (2023). Penerapan Metode Regresi Linear untuk Prediksi Penjualan Suku Cadang. *Jurnal Informatika Teknologi dan Sains (Jinteks)*, 5(4), 679–685. DOI: <https://doi.org/10.51401/jinteks.v5i4.3462> .
- [21]Florensa Nainggolan, N. C., Boy, A. F., & Elfitriani, E. (2023). Penerapan Data Mining Untuk Prediksi Export Penjualan Produk Kerajinan Rotan Menggunakan Metode Regresi Linear Berganda. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGД)*, 2(5), 743. DOI: <https://doi.org/10.53513/jursi.v2i5.6779> .
- [22]Novita Indriyani, Heru Satria Tambunan, & Zulia Almaida Siregar. (2022). Analisis Faktor Kepuasan Konsumen Terhadap Produk Roti Pinkan Bakery & Cake dengan Algoritma C4.5. *Jurnal Riset Rumpun Ilmu Teknik*, 1(2), 76–90. DOI: <https://doi.org/10.55606/jurritek.v1i2.413> .