**Jurnal Pustaka Cendekia Hukum dan Ilmu Sosial**
Volume 4, No. 2, June-September 2026, pp. 3301-3313
E-ISSN: 2987-4475
https://pchukumsosial.org/index.php/pchs

# Klasifikasi Status Stok Barang Distribusi untuk Persediaan Menggunakan Algoritma Naive Bayes Berdasarkan Histori Data Penjualan PT. Berlian Djaya Nusantara

Febri Dwi Hariyanto[1], Imam Budiawan[2*], Yumi Novita Dewi[3]

[1,2,3]Informatika, Universitas Bina Sarana Informatika, Jl. Kramat Raya No.98, Kec. Senen, Kota Jakarta Pusat, DKI Jakarta
imam.imb@bsi.ac.id

***Abstract***

Inventory management in distribution companies is exposed to excess-stock and shortage risks when replenishment decisions rely mainly on estimates. The sales history of PT Berlian Djaya Nusantara had not been systematically used to distinguish the stock condition of each item. The analysis classified stock status using the Naive Bayes algorithm and 250 sales records from 2025. The processing stages covered data selection, quality checking, transformation, target-class construction, modeling, and evaluation using a confusion matrix. Stock status was derived from Ending Stock through quartile-based discretization, producing Q1 of 138 and Q3 of 311. These thresholds generated 63 Low, 124 Medium, and 63 High records. Sales Quantity, Sales Frequency, and Initial Stock served as predictors, while Stock Status became the classification target. Five training-to-testing ratios were examined: 90:10, 80:20, 70:30, 60:40, and 50:50. The 90:10 ratio produced the strongest performance, with 70.83% accuracy, 70.27% precision, 72.22% recall, and a 71.23% F1-score. The model distinguished the three stock categories and generated measurable information for inventory control, although classification errors remained because the relationships among several attributes were not fully represented by the model.
**Keywords:** Data Mining, Sales History, Classification, Naive Bayes, Stock Status

**Abstrak**

Pengelolaan persediaan pada perusahaan distribusi menghadapi risiko penumpukan dan kekurangan barang ketika keputusan stok hanya bertumpu pada perkiraan. Histori penjualan PT Berlian Djaya Nusantara belum dimanfaatkan secara terstruktur untuk membedakan kondisi stok setiap barang. Analisis bertujuan mengklasifikasikan status stok dengan algoritma Naive Bayes berdasarkan 250 catatan penjualan tahun 2025. Tahapan pengolahan mencakup pemilihan data, pemeriksaan kualitas, transformasi, pembentukan kelas target, pemodelan, serta evaluasi menggunakan confusion matrix. Status stok dibentuk dari nilai Stok Akhir melalui metode kuartil dengan Q1 sebesar 138 dan Q3 sebesar 311. Batas tersebut menghasilkan 63 data berkategori Rendah, 124 data Sedang, dan 63 data Tinggi. Atribut yang digunakan sebagai prediktor terdiri atas Jumlah Penjualan, Frekuensi Penjualan, dan Stok Awal, sedangkan Status Stok menjadi target klasifikasi. Pengujian dilakukan melalui lima pembagian data latih dan data uji, yaitu 90:10, 80:20, 70:30, 60:40, dan 50:50. Rasio 90:10 menghasilkan kinerja terbaik dengan accuracy 70,83%, precision 70,27%, recall 72,22%, dan F1-score 71,23%. Model mampu membedakan tiga kategori stok dan memberi informasi terukur bagi pengendalian persediaan, walaupun kesalahan klasifikasi masih muncul akibat pola antaratribut yang belum sepenuhnya terwakili.
**Kata Kunci:** Data Mining, Histori Penjualan, Klasifikasi, Naive Bayes, Status Stok



✉ Corresponding author: Febri Dwi Hariyanto
Email Address: febrihriyanto@gmail.com (Jl. Kramat Raya No.98, Kec. Senen, Kota Jakarta Pusat, DKI Jakarta)
Received 01 August 2026, Accepted 13 August 2026, Published 26 August 2026

## PENDAHULUAN

Pengelolaan persediaan menjadi bagian pokok dari kegiatan distribusi karena ketersediaan barang menentukan kelancaran pelayanan, biaya penyimpanan, serta kemampuan perusahaan memenuhi permintaan pelanggan. Perubahan volume penjualan dari satu periode ke periode lain membuat jumlah stok tidak dapat ditetapkan hanya melalui kebiasaan operasional. Persediaan yang terlalu besar menahan modal, menambah kebutuhan ruang gudang, dan meningkatkan risiko kerusakan, sedangkan persediaan yang terlalu kecil menghambat pemenuhan pesanan. Kondisi tersebut menuntut

keputusan yang lebih terukur agar jumlah barang yang disimpan sejalan dengan pola permintaan aktual. Data transaksi yang terus terbentuk selama kegiatan penjualan dapat memberi gambaran mengenai pergerakan barang apabila diolah secara sistematis. Keputusan berbasis data juga mengurangi ketergantungan pada penilaian subjektif petugas gudang atau pihak manajemen. Keteraturan pengendalian stok akhirnya berkaitan langsung dengan efisiensi operasional perusahaan distribusi (Manajemen & Sulistyawati, 2024).

Persediaan barang jadi pada perusahaan distribusi berfungsi sebagai penghubung antara pengadaan dan kebutuhan pelanggan. Keseimbangan jumlah stok perlu dipertahankan karena permintaan dapat berubah menurut jenis barang, intensitas transaksi, dan periode penjualan. Penumpukan barang memunculkan biaya pemeliharaan serta penggunaan kapasitas gudang yang tidak efisien, sedangkan kekurangan barang menimbulkan kehilangan peluang penjualan. Pengendalian persediaan juga harus memperhatikan kondisi awal stok, jumlah barang yang terjual, dan sisa barang setelah transaksi berlangsung. Ketiga informasi tersebut tidak berdiri sendiri karena perubahan satu unsur akan memengaruhi kondisi persediaan pada akhir periode. Catatan historis yang tersusun baik memberi dasar untuk menilai apakah posisi stok berada pada tingkat rendah, sedang, atau tinggi. Pemahaman tersebut mendukung perencanaan pengadaan tanpa memisahkan kebutuhan pelanggan dari kemampuan penyimpanan perusahaan (Heizer et al., 2022).

PT Berlian Djaya Nusantara menjalankan kegiatan distribusi berbagai barang penunjang konstruksi dan menghasilkan catatan penjualan secara berulang. Data perusahaan memuat kode barang, nama barang, kategori produk, periode transaksi, jumlah penjualan, satuan, frekuensi penjualan, stok awal, dan stok akhir. Catatan tersebut sebelumnya lebih banyak berfungsi sebagai arsip transaksi daripada sebagai dasar klasifikasi persediaan. Penentuan stok yang masih mengandalkan pengalaman dapat memunculkan perbedaan antara jumlah barang yang tersedia dan kebutuhan pasar. Risiko kekurangan stok terlihat ketika pesanan tidak dapat dipenuhi pada waktu yang dibutuhkan, sedangkan stok berlebih memperbesar beban penyimpanan. Pengelompokan kondisi stok dibutuhkan agar perusahaan dapat melihat posisi setiap catatan penjualan secara lebih objektif. Masalah ketersediaan barang di gudang juga berkaitan dengan tingkat layanan karena barang yang tidak tersedia dapat memperlambat pemenuhan permintaan pelanggan (Pahlevi, 2026).

Pemanfaatan data mining memungkinkan histori penjualan diubah menjadi pola yang mendukung pengambilan keputusan. Proses tersebut menggabungkan pengelolaan basis data, statistika, dan teknik pembelajaran mesin untuk menemukan keteraturan yang sulit diamati melalui pemeriksaan manual. Klasifikasi menjadi salah satu fungsi utama karena setiap catatan dapat ditempatkan ke dalam kelas yang sudah ditentukan. Model dibangun dari data berlabel, mempelajari hubungan antara atribut prediktor dan target, lalu digunakan untuk memperkirakan kelas data uji. Algoritma Naive Bayes bekerja melalui probabilitas dengan asumsi bahwa atribut memberikan kontribusi secara independen terhadap kelas. Kesederhanaan perhitungan membuat algoritma tersebut dapat diterapkan pada data numerik maupun kategorikal dengan kebutuhan komputasi yang relatif rendah. Kerangka tersebut

memberi landasan bagi pengolahan histori penjualan menjadi informasi status stok yang dapat diukur (Han et al., 2022).

Sejumlah kajian terdahulu memperlihatkan pemanfaatan metode klasifikasi untuk membaca pola penjualan dan persediaan. Andriansyah (2021) melaporkan bahwa Naive Bayes dapat mengelompokkan penjualan barang pada perusahaan ritel berdasarkan transaksi yang tersedia. Pratama dan Nugroho (2022) menempatkan histori penjualan sebagai dasar pengurangan risiko overstock dan stockout pada perusahaan distribusi. Wijaya (2023) menggunakan data historis untuk mengklasifikasikan tingkat permintaan sehingga pengelolaan persediaan menjadi lebih terarah. Hidayat (2024) menunjukkan bahwa pola permintaan yang diperoleh melalui data mining dapat membantu penyusunan strategi pengadaan dan distribusi. Empat kajian tersebut mempunyai kesamaan pada penggunaan transaksi masa lalu, tetapi objek, atribut, pembentukan kelas, dan skenario evaluasinya berbeda. Perbedaan itu membuka ruang untuk menguji klasifikasi tiga tingkat status stok dengan batas kelas yang berasal dari distribusi data perusahaan.

Celah analisis terletak pada pembentukan target stok yang tidak ditentukan melalui perkiraan manajerial, melainkan melalui kuartil Stok Akhir dari seluruh data. Pendekatan tersebut menghasilkan batas kelas yang mengikuti sebaran aktual dan mengurangi penetapan kategori secara subjektif. Stok Akhir hanya digunakan untuk membentuk target dan dikeluarkan dari prediktor agar model tidak memperoleh jawaban secara langsung dari atribut pembentuk kelas. Jumlah Penjualan, Frekuensi Penjualan, dan Stok Awal kemudian berfungsi sebagai informasi yang dipelajari oleh algoritma. Lima rasio pembagian data digunakan untuk melihat perubahan kinerja ketika proporsi data latih dikurangi secara bertahap. Rancangan tersebut tidak sekadar menghasilkan satu nilai akurasi, tetapi memperlihatkan kestabilan model pada komposisi data yang berbeda. Pengolahan data penjualan secara sistematis juga memberi dasar rekomendasi stok yang lebih terukur daripada prosedur konvensional (Ramadhan & Putri, 2023).

Tujuan analisis diarahkan pada penerapan algoritma Naive Bayes, pemanfaatan histori penjualan sebagai dasar pembentukan status stok, pengukuran kinerja model, serta penafsiran hasil klasifikasi bagi pengendalian persediaan. Kontribusi utama berada pada penggunaan kuartil untuk membentuk kelas Rendah, Sedang, dan Tinggi, pemisahan atribut pembentuk target dari prediktor, serta perbandingan lima rasio data latih dan data uji. Hasil akhir diharapkan memberi gambaran terukur mengenai kemampuan model mengenali posisi stok melalui atribut penjualan yang tersedia. Informasi kelas dapat digunakan sebagai bahan evaluasi internal ketika perusahaan meninjau barang dengan sisa stok terlalu rendah atau terlalu tinggi. Batas kontribusi tetap ditempatkan pada klasifikasi status, bukan pada peramalan jumlah kebutuhan atau penetapan otomatis kuantitas pembelian. Pemisahan tersebut menjaga agar simpulan sesuai dengan data dan proses analisis yang benar-benar dilakukan. Pemanfaatan histori transaksi untuk pengelolaan persediaan telah dipandang mampu menghasilkan informasi operasional yang lebih objektif (Julianto & Andayani, 2024).

## METODE

### *Kerangka Dasar Penelitian*

Rancangan analisis menggunakan pendekatan kuantitatif berbasis data mining dengan tugas utama klasifikasi. Data numerik berasal dari histori penjualan PT Berlian Djaya Nusantara dan diproses melalui tahapan Knowledge Discovery in Database. Urutan kerja mencakup pemilihan atribut, pemeriksaan kualitas, pembentukan kelas target, transformasi format, pembagian data, pemodelan, dan evaluasi. Penggunaan data historis bertujuan menemukan hubungan antara aktivitas penjualan dan posisi stok yang tersisa. Hasil pemodelan tidak dipakai untuk memperkirakan nilai Stok Akhir, melainkan untuk menentukan kelas status stok. Pengelompokan tersebut mengikuti skema pembelajaran terawasi karena setiap catatan telah mempunyai target setelah proses diskretisasi. Pendekatan berbasis data mining membantu mengekstraksi informasi yang sebelumnya tersembunyi di dalam kumpulan transaksi penjualan (Sari & Nasution, 2025).

Objek analisis berupa catatan distribusi barang selama tahun 2025 yang tersimpan dalam format lembar data perusahaan. Sumber primer diperoleh melalui dokumentasi histori penjualan, sedangkan observasi dan wawancara memberi gambaran mengenai pengelolaan stok yang berjalan. Data sekunder pada naskah asal terdiri atas buku dan artikel yang membahas data mining, persediaan, klasifikasi, serta evaluasi model. Unit analisis berada pada setiap catatan barang dan periode transaksi, bukan pada pelanggan atau pemasok. Proses klasifikasi memerlukan pemisahan data latih dan data uji agar kemampuan model dapat diuji pada catatan yang tidak digunakan saat pembentukan probabilitas. Model probabilistik memilih kelas dengan nilai posterior tertinggi setelah memperhitungkan probabilitas awal kelas dan peluang setiap atribut. Mekanisme tersebut mengikuti prinsip umum klasifikasi terawasi yang membedakan tahap pembelajaran dan tahap pengujian (Han et al., 2022).

### *Sumber Data dan Variabel*

Dataset terdiri atas 250 catatan penjualan dengan atribut identitas, transaksi, dan persediaan. Pemeriksaan meliputi data kosong, duplikasi, nilai negatif, kesesuaian satuan, dan konsistensi kategori. Seluruh data lengkap dan tidak ditemukan duplikasi, sehingga tidak ada baris yang dihapus atau diimputasi. Kode Barang dan Nama Barang dipertahankan sebagai identitas, tetapi tidak digunakan dalam model. Stok Akhir digunakan untuk membentuk target Status Stok, kemudian dikeluarkan dari prediktor untuk mencegah data leakage. Tahap ini memastikan performa model tidak dipengaruhi kesalahan input dan mendukung ketepatan klasifikasi (Lubis & Sitohang, 2023).

Variabel akhir terdiri atas tiga prediktor numerik dan satu target nominal. Jumlah Penjualan menunjukkan jumlah barang terjual, Frekuensi Penjualan menunjukkan jumlah transaksi, dan Stok Awal mencerminkan persediaan sebelum penjualan. Status Stok menjadi target yang diklasifikasikan menjadi Rendah, Sedang, atau Tinggi. Ketiga prediktor dipilih berdasarkan informasi yang tersedia sebelum status akhir diketahui, sehingga kondisi persediaan dapat dinilai tanpa menggunakan Stok Akhir sebagai masukan. Jenis data disesuaikan agar atribut numerik terbaca sebagai bilangan dan target

sebagai kelas. Struktur variabel ini sesuai dengan fungsi persediaan sebagai penyangga antara pengadaan dan permintaan (Heizer et al., 2022).

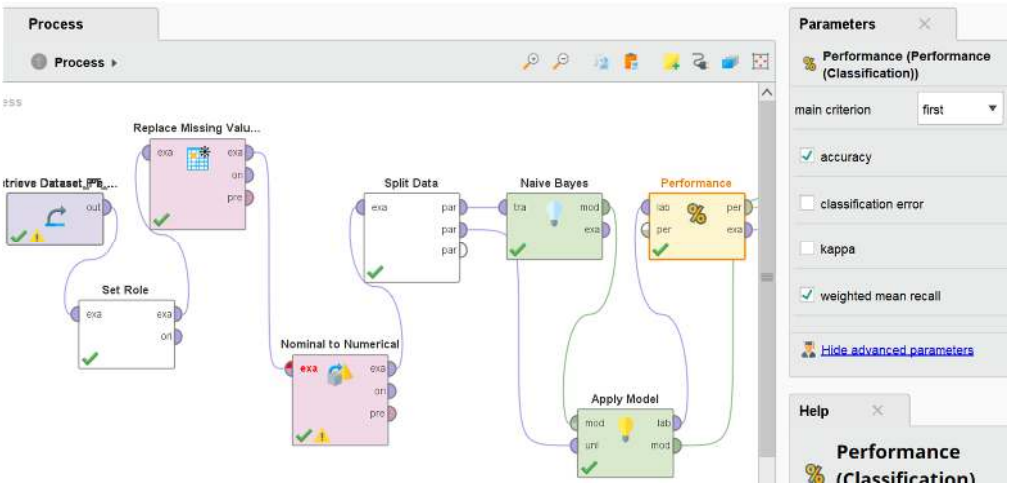
Tabel 1. Variabel Pemodelan

| No | Variabel | Tipe Data | Peran |
|---|---|---|---|
| 1 | Jumlah Penjualan | Integer | Prediktor |
| 2 | Frekuensi Penjualan | Integer | Prediktor |
| 3 | Stok Awal | Integer | Prediktor |
| 4 | Status Stok | Nominal | Target |

***Tahapan Penelitian***

Tahapan pengolahan diawali dengan memilih 250 catatan penjualan yang memiliki atribut relevan terhadap persediaan. Data diperiksa untuk memastikan tidak terdapat nilai kosong, duplikasi, atau format tidak konsisten. Stok Akhir diurutkan untuk menentukan kuartil pertama dan ketiga sebagai batas pembentukan target, kemudian setiap catatan diberi label Rendah, Sedang, atau Tinggi berdasarkan rentang kuartil. Seleksi atribut menghapus informasi identitas dan Stok Akhir, sementara tiga prediktor numerik dipertahankan. Dataset selanjutnya dibagi menjadi data latih dan uji dengan lima rasio. Tahapan ini mengubah data historis menjadi pola yang mendukung keputusan persediaan (Aditiya et al., 2020).

Pemodelan dilakukan menggunakan RapidMiner Studio 2026 melalui operator pengambilan data, penetapan target, pemeriksaan nilai kosong, transformasi, pembagian data, Naive Bayes, penerapan model, dan evaluasi klasifikasi. Status Stok ditetapkan sebagai label, sedangkan data latih digunakan untuk membentuk probabilitas kelas dan data uji diproses melalui Apply Model. Prediksi kemudian dibandingkan dengan kelas aktual menggunakan Performance Classification dengan metrik accuracy, precision, recall, dan F1-score. Rangkaian operator menghasilkan alur yang dapat ditelusuri dari data awal hingga evaluasi. Komputasi Naive Bayes yang sederhana mendukung klasifikasi data historis secara efisien (Saepudin et al., 2023).




Gambar 1. Rangkaian Pemodelan Naive Bayes pada RapidMiner
Sumber: Dokumentasi pengolahan data skripsi, 2026

***Pembentukan Kategori Status Stok***

Pembentukan target dilakukan dengan metode kuartil terhadap 250 nilai Stok Akhir yang telah diurutkan dari nilai terendah hingga tertinggi. Posisi Q1 berada pada data ke-63 sampai ke-64, dengan nilai 137 dan 141, sehingga interpolasi menghasilkan Q1 sebesar 138. Posisi Q3 berada pada data ke-187 sampai ke-188, dengan nilai 308 dan 312, sehingga interpolasi menghasilkan Q3 sebesar 311. Catatan dengan Stok Akhir di bawah 138 diberi kelas Rendah. Nilai antara 138 dan 311, termasuk kedua batas, ditempatkan pada kelas Sedang. Nilai di atas 311 ditempatkan pada kelas Tinggi. Diskretisasi berbasis kuartil membentuk kelas dari distribusi aktual dan mengurangi penggunaan batas yang ditetapkan secara subjektif (Giovanno & Widyadana, 2022).

Tabel 2. Kriteria Status Stok

| **Nilai** Stok Akhir | Status Stok |
|---|---|
| Stok Akhir < 138 | Rendah |
| 138 ≤ Stok Akhir ≤ 311 | Sedang |
| Stok Akhir > 311 | Tinggi |

Kelas yang sudah terbentuk berfungsi sebagai target pada proses pembelajaran, sedangkan Stok Akhir tidak lagi masuk ke model. Pemisahan tersebut diperlukan karena target berasal langsung dari nilai Stok Akhir. Penggunaan atribut yang sama sebagai target dan prediktor akan membuat model membaca informasi yang seharusnya hanya tersedia setelah klasifikasi. Kondisi itu dapat menghasilkan performa semu yang tidak menggambarkan kemampuan algoritma mengenali pola dari atribut penjualan. Seleksi variabel memastikan bahwa prediksi dibentuk melalui Jumlah Penjualan, Frekuensi Penjualan, dan Stok Awal. Keputusan tersebut juga menjaga hubungan logis antara masukan dan keluaran pemodelan. Pencegahan kebocoran informasi menjadi bagian dari pengendalian kualitas data sebelum evaluasi dilakukan (Lubis & Sitohang, 2023).

***Pemodelan dan Evaluasi***

Pembagian data menggunakan lima komposisi agar perubahan performa dapat diamati ketika jumlah data latih dan data uji berubah. Rasio 90:10 menghasilkan 225 data latih dan 25 data uji, sedangkan rasio 50:50 membagi dataset menjadi 125 data latih dan 125 data uji. Tiga komposisi lain terdiri atas 200:50, 175:75, dan 150:100. Distribusi kelas dipertahankan pada setiap bagian melalui stratified sampling sehingga proporsi Rendah, Sedang, dan Tinggi tidak berubah secara tajam. Setiap skenario menjalani alur pemodelan dan evaluasi yang sama. Perbedaan hasil kemudian dikaitkan dengan banyaknya informasi yang tersedia bagi algoritma saat mempelajari probabilitas kelas. Pembagian data yang terkontrol diperlukan untuk menilai kemampuan generalisasi model, bukan sekadar kecocokannya terhadap data latih (Saepudin et al., 2023).

Tabel 3. Skenario Pembagian Data

| **Rasio Latih:Uji** | **Data Latih** | **Data Uji** | **Total** |
|---|---|---|---|
| 90:10 | 225 | 25 | 250 |
| 80:20 | 200 | 50 | 250 |

| Rasio Latih:Uji | Data Latih | Data Uji | Total |
|---|---|---|---|
| 70:30 | 175 | 75 | 250 |
| 60:40 | 150 | 100 | 250 |
| 50:50 | 125 | 125 | 250 |

Evaluasi menggunakan confusion matrix multikelas berukuran 3 × 3 dengan kelas Rendah, Sedang, dan Tinggi. Nilai diagonal menunjukkan jumlah prediksi yang sesuai dengan kelas aktual, sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi. Perhitungan precision, recall, dan F1-score dilakukan melalui pendekatan One-vs-Rest untuk setiap kelas, lalu diringkas sebagai ukuran kinerja model. Accuracy memperlihatkan proporsi seluruh data uji yang diprediksi dengan benar. Precision menggambarkan ketepatan prediksi suatu kelas, sedangkan recall menunjukkan kemampuan menemukan anggota kelas yang sebenarnya. F1-score menyeimbangkan kedua ukuran tersebut ketika salah satu nilai tidak cukup mewakili performa. Rangkaian metrik memberikan penilaian yang lebih lengkap daripada penggunaan akurasi tunggal (Han et al., 2022).

## HASIL DAN DISKUSI

### *Kualitas dan Transformasi Data*

Dataset awal terdiri atas 250 catatan dan seluruhnya dapat dipertahankan setelah pemeriksaan kualitas. Kolom yang diperiksa meliputi kode barang, nama barang, kategori produk, periode transaksi, jumlah penjualan, satuan, frekuensi penjualan, stok awal, dan stok akhir. Tidak ditemukan nilai kosong maupun catatan ganda, sehingga jumlah data sebelum dan sesudah pembersihan tetap sama. Nilai Jumlah Penjualan dan Frekuensi Penjualan juga tidak menunjukkan angka negatif. Keseragaman satuan dan kategori barang diperiksa agar variasi penulisan tidak membentuk kelompok yang sebenarnya sama. Operator Replace Missing Values tetap disertakan pada alur perangkat lunak sebagai langkah pengamanan, walaupun kolom kehilangan data menunjukkan nilai nol. Hasil tersebut menegaskan bahwa kualitas masukan telah dikendalikan sebelum model mempelajari pola klasifikasi (Lubis & Sitohang, 2023).

Transformasi data menghasilkan struktur akhir yang lebih ringkas daripada dataset awal. Kode Barang dan Nama Barang tidak digunakan karena hanya berfungsi sebagai identitas, sedangkan Stok Akhir dikeluarkan karena membentuk target. Jumlah Penjualan, Frekuensi Penjualan, dan Stok Awal disimpan sebagai bilangan bulat. Status Stok disimpan sebagai data nominal agar tiga kelas dapat dibedakan oleh operator klasifikasi. Seleksi tersebut mengurangi atribut yang tidak memberi informasi langsung kepada model dan menjaga agar hubungan prediksi tidak tercampur dengan identitas produk. Dataset akhir tetap mempertahankan seluruh 250 baris karena seleksi dilakukan pada kolom, bukan pada catatan. Pengolahan histori penjualan yang terstruktur dapat mengubah arsip transaksi menjadi informasi persediaan yang lebih mudah dianalisis (Julianto & Andayani, 2024).

***Pembentukan Kategori Status Stok***

Nilai Q1 sebesar 138 dan Q3 sebesar 311 membagi data ke dalam tiga kelompok dengan pola 25%, 50%, dan 25% secara mendekati. Kelas Rendah berisi 63 catatan atau 25,2% dari total data. Kelas Sedang berisi 124 catatan atau 49,6%, sedangkan kelas Tinggi kembali berisi 63 catatan atau 25,2%. Komposisi tersebut muncul sebagai akibat langsung dari pemotongan kuartil dan bukan dari penyeimbangan buatan setelah target terbentuk. Kelas Sedang mempunyai anggota lebih banyak karena menampung seluruh nilai antara Q1 dan Q3. Dua kelas ekstrem mempunyai jumlah yang sama dan masing-masing mewakili bagian bawah serta bagian atas distribusi stok akhir. Penggunaan kuartil memberi dasar kategorisasi yang mengikuti susunan data perusahaan (Giovanno & Widyadana, 2022).

Tabel 4. Distribusi Kelas Status Stok

| Kategori | Jumlah Data | Persentase |
|---|---|---|
| Rendah | 63 | 25,2% |
| Sedang | 124 | 49,6% |
| Tinggi | 63 | 25,2% |
| Total | 250 | 100% |

Distribusi kelas memperlihatkan ketidakseimbangan ringan karena kelas Sedang hampir dua kali lebih besar daripada setiap kelas ekstrem. Keadaan tersebut masih dapat dijelaskan oleh metode kuartil karena separuh bagian tengah memang ditempatkan pada satu kategori. Pemakaian stratified sampling diperlukan agar kelas Rendah dan Tinggi tidak semakin berkurang secara tidak proporsional pada data uji. Kelas dengan anggota lebih banyak cenderung memberi contoh yang lebih beragam kepada model, sedangkan kelas yang lebih kecil mempunyai variasi contoh terbatas. Perbedaan tersebut dapat memengaruhi jumlah prediksi benar pada masing-masing kelas. Kondisi distribusi perlu dibaca bersama precision dan recall agar performa tidak disimpulkan hanya melalui akurasi total. Pemanfaatan histori penjualan untuk mengurangi risiko stok membutuhkan perhatian terhadap pola permintaan dan komposisi data yang dianalisis (Pratama & Nugroho, 2022).

***Pembagian Data dan Implementasi Model***

Lima skenario menghasilkan jumlah contoh pembelajaran yang berbeda bagi algoritma. Rasio 90:10 menyediakan 225 data latih dan hanya 25 data uji, sehingga model memperoleh bagian terbesar dari dataset untuk membentuk probabilitas. Rasio 80:20 mengurangi data latih menjadi 200, sedangkan rasio 70:30 dan 60:40 masing-masing menggunakan 175 serta 150 data latih. Rasio 50:50 memberikan jumlah data latih paling sedikit, yaitu 125 catatan. Setiap pembagian tetap menggunakan atribut dan parameter evaluasi yang sama. Perbandingan tersebut memungkinkan perubahan kinerja dikaitkan dengan komposisi data, bukan dengan perubahan metode atau target. Pemilihan Naive Bayes bertumpu pada proses komputasi yang sederhana dan kemampuan mengolah data historis dengan cepat (Saepudin et al., 2023).

Implementasi pada RapidMiner Studio 2026 dimulai dari pengambilan dataset dan pengaturan Status Stok sebagai target. Operator pembersihan memastikan tidak terdapat nilai kosong sebelum data

memasuki proses pemodelan. Data yang telah ditransformasi kemudian dibagi sesuai rasio, dengan satu keluaran digunakan untuk pelatihan dan keluaran lain dipakai untuk pengujian. Operator Naive Bayes membentuk model berdasarkan probabilitas kelas Rendah, Sedang, dan Tinggi. Apply Model menghasilkan prediksi pada data uji, lalu Performance Classification membandingkannya dengan kelas aktual. Alur yang sama dijalankan berulang pada seluruh skenario agar hasil dapat dibandingkan secara langsung. Penerapan data mining pada manajemen persediaan memberi cara untuk membaca pola permintaan yang tidak mudah terlihat dari tabel transaksi mentah (Hidayat, 2024).

***Hasil Evaluasi Model***

Hasil pengujian menunjukkan variasi kinerja yang cukup lebar antarproporsi data. Rasio 90:10 menghasilkan accuracy tertinggi sebesar 70,83%, sedangkan rasio 50:50 menghasilkan nilai terendah sebesar 55,20%. Rasio 70:30 mencapai 64,00% dan sedikit lebih baik daripada rasio 60:40 sebesar 62,00%. Rasio 80:20 hanya mencapai 58,82%, sehingga kenaikan jumlah data latih tidak selalu menghasilkan urutan performa yang sepenuhnya linear. Variasi tersebut dapat dipengaruhi komposisi contoh yang terpilih pada setiap pembagian, walaupun proporsi kelas dipertahankan. Model tetap mampu memprediksi ketiga kelas pada seluruh skenario, tetapi jumlah kesalahan meningkat ketika data latih berkurang. Kemampuan Naive Bayes mengelompokkan transaksi penjualan juga pernah terlihat pada klasifikasi barang ritel (Andriansyah, 2021).

Tabel 5. Perbandingan Performa Model

| Rasio Latih:Uji | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| 90:10 | 70,83% | 70,27% | 72,22% | 71,23% |
| 80:20 | 58,82% | 59,96% | 57,23% | 58,56% |
| 70:30 | 64,00% | 63,89% | 62,02% | 62,94% |
| 60:40 | 62,00% | 61,80% | 61,33% | 61,56% |
| 50:50 | 55,20% | 58,52% | 51,71% | 54,90% |

Rasio 90:10 menjadi skenario terbaik pada seluruh ukuran evaluasi. Sebanyak 17 dari 24 catatan yang terbaca pada ringkasan confusion matrix diprediksi benar, terdiri atas lima kelas Rendah, delapan kelas Sedang, dan empat kelas Tinggi, sementara tujuh catatan salah klasifikasi. Nilai precision 70,27% menunjukkan bahwa sebagian besar kelas yang diprediksi model sesuai dengan kondisi aktual. Nilai recall 72,22% menjadi ukuran tertinggi pada skenario tersebut dan menunjukkan kemampuan model menemukan anggota kelas aktual. F1-score 71,23% memperlihatkan keseimbangan yang relatif baik antara ketepatan dan jangkauan prediksi. Proporsi data latih yang besar memberi lebih banyak contoh bagi pembentukan probabilitas setiap kelas. Pemanfaatan data historis melalui Naive Bayes juga dilaporkan membantu klasifikasi tingkat permintaan pada sistem persediaan (Wijaya, 2023).

Empat skenario lain memperlihatkan penurunan kinerja dengan pola yang tidak seragam. Rasio 80:20 mempunyai accuracy 58,82% dan 21 kesalahan klasifikasi, sedangkan rasio 70:30 meningkat menjadi 64,00% walaupun jumlah data latih lebih sedikit. Rasio 60:40 menghasilkan 38 kesalahan dari 100 data uji dan nilai F1-score 61,56%. Rasio 50:50 menghasilkan 56 kesalahan dari 125 data uji serta

recall 51,71%, yang menjadi nilai jangkauan prediksi terendah. Perubahan tersebut menunjukkan bahwa jumlah data latih bukan satu-satunya faktor karena susunan contoh pada setiap bagian juga memengaruhi estimasi probabilitas. Kinerja yang naik pada 70:30 setelah turun pada 80:20 memperlihatkan adanya variasi sampel yang perlu diperhatikan ketika hasil ditafsirkan. Pengolahan data penjualan yang sistematis tetap memberikan dasar klasifikasi yang lebih terukur daripada keputusan tanpa analisis data (Ramadhan & Putri, 2023).

Kategori Sedang menjadi kelas yang paling konsisten dikenali karena jumlah prediksi benar lebih tinggi pada hampir seluruh skenario. Jumlah tersebut mencapai delapan pada rasio 90:10, enam belas pada 80:20, dua puluh enam pada 70:30, tiga puluh dua pada 60:40, dan empat puluh satu pada 50:50. Keunggulan tersebut berkaitan dengan jumlah anggota kelas Sedang yang mencapai 124 data, hampir dua kali setiap kelas lain. Kategori Tinggi menunjukkan prediksi benar yang lebih rendah pada beberapa skenario, terutama ketika data uji bertambah. Kelas Rendah juga mengalami perubahan jumlah prediksi benar sesuai ukuran data uji. Perbandingan per kelas menjelaskan mengapa akurasi total perlu dibaca bersama distribusi kategori. Penggunaan klasifikasi untuk mengurangi risiko overstock dan stockout memerlukan perhatian terhadap kemampuan model mengenali kelas ekstrem, bukan hanya kelas mayoritas (Pratama & Nugroho, 2022).

***Pembahasan***

Temuan utama memperlihatkan bahwa porsi data latih yang lebih besar memberi keuntungan pada skenario 90:10, tetapi hubungan antara proporsi dan kinerja tidak sepenuhnya lurus. Algoritma probabilistik memerlukan contoh yang cukup untuk memperkirakan distribusi setiap atribut pada masing-masing kelas. Jumlah data latih yang berkurang membatasi variasi pola yang dapat dipelajari, terutama pada kelas Rendah dan Tinggi yang masing-masing hanya mempunyai 63 catatan. Kenaikan performa dari 80:20 ke 70:30 menunjukkan bahwa komposisi baris dalam data uji turut berperan. Satu kali pembagian pada setiap rasio belum menggambarkan seluruh kemungkinan variasi sampel. Hasil tetap mendukung penggunaan rasio 90:10 sebagai model terbaik di antara skenario yang benar-benar diuji. Prinsip pemisahan data latih dan data uji memang ditujukan untuk menilai kemampuan model menghadapi catatan yang tidak dipakai saat pembelajaran (Han et al., 2022).

Perbandingan dengan kajian sebelumnya menunjukkan kesamaan pada kemampuan Naive Bayes membaca pola transaksi, tetapi nilai performa tidak dapat disamakan secara langsung karena struktur data dan pembentukan target berbeda. Andriansyah (2021) menempatkan algoritma pada klasifikasi penjualan barang ritel dan memperoleh hasil yang dinilai tinggi pada konteks datanya. Wijaya (2023) memanfaatkan data historis penjualan untuk klasifikasi tingkat permintaan pada sistem persediaan. Analisis PT Berlian Djaya Nusantara memakai target yang dibentuk dari kuartil Stok Akhir dan hanya menggunakan tiga prediktor numerik pada model akhir. Perbedaan tersebut menghasilkan tugas klasifikasi yang lebih spesifik terhadap posisi stok, bukan sekadar tingkat penjualan. Nilai accuracy 70,83% menunjukkan kemampuan yang cukup, tetapi belum mendukung klaim bahwa model telah

mencapai ketepatan tinggi. Pembacaan yang hati-hati menjaga agar perbandingan tidak mengabaikan perbedaan dataset, kelas, dan prosedur evaluasi.

Implikasi operasional terletak pada penyediaan informasi kategori yang dapat dipakai untuk meninjau catatan stok secara lebih terarah. Kelas Rendah dapat menjadi sinyal awal bagi perusahaan untuk memeriksa kemungkinan kekurangan persediaan, sedangkan kelas Tinggi menunjukkan perlunya evaluasi terhadap penumpukan barang. Kelas Sedang menggambarkan posisi stok di antara dua kuartil dan tidak otomatis berarti jumlah tersebut selalu ideal untuk setiap jenis produk. Hasil klasifikasi perlu dibaca bersama informasi pengadaan, kecepatan distribusi, dan kebutuhan aktual pelanggan sebelum keputusan pembelian dibuat. Model tidak menghasilkan kuantitas pemesanan maupun waktu pemesanan kembali, sehingga fungsinya berada pada penyediaan kategori. Batas tersebut penting agar keluaran klasifikasi tidak diperlakukan sebagai keputusan persediaan yang sepenuhnya otomatis. Pengelolaan ketidaktersediaan barang di gudang tetap membutuhkan strategi layanan dan pengadaan yang mempertimbangkan kondisi operasional perusahaan (Pahlevi, 2026).

Kesalahan klasifikasi dapat dijelaskan melalui asumsi independensi yang menjadi dasar Naive Bayes. Jumlah Penjualan, Frekuensi Penjualan, dan Stok Awal secara operasional mungkin mempunyai hubungan yang tidak sepenuhnya terpisah. Jumlah barang yang terjual dapat berkaitan dengan frekuensi transaksi, sedangkan keduanya berhubungan dengan perubahan dari stok awal menuju stok akhir. Asumsi independensi membuat interaksi antarprediktor tidak dipelajari secara eksplisit. Jumlah atribut yang terbatas juga belum menangkap faktor seperti stok minimum, stok maksimum, jumlah pembelian, pemasok, musim, lead time, dan reorder point. Keterbatasan tersebut memberi penjelasan mengapa sejumlah catatan tetap masuk ke kelas yang salah. Kinerja Naive Bayes dapat menurun ketika terdapat korelasi kuat antaratribut atau pola hubungan yang lebih kompleks (Saepudin et al., 2023).

Kontribusi analisis tidak hanya berada pada nilai evaluasi, tetapi juga pada cara target dibentuk dan kebocoran informasi dihindari. Penggunaan kuartil menghasilkan batas 138 dan 311 yang dapat ditelusuri dari susunan data, sedangkan penghapusan Stok Akhir dari prediktor menjaga model tetap fair. Lima rasio data memberi gambaran bahwa performa dapat berubah ketika komposisi pembelajaran dan pengujian digeser. Rangkaian tersebut menyediakan dasar yang lebih transparan daripada penetapan kelas tanpa aturan distribusi. Kebaruan praktis muncul dari penerapan skema tersebut pada histori penjualan PT Berlian Djaya Nusantara. Nilai terbaik tetap perlu ditempatkan sebagai hasil dari dataset dan konfigurasi yang tersedia, bukan sebagai performa yang pasti sama pada periode lain. Analisis pola permintaan melalui data mining mendukung penyusunan strategi persediaan ketika hasilnya ditafsirkan sesuai batas data (Hidayat, 2024).

Hipotesis kerja yang menyatakan bahwa Naive Bayes mampu mengklasifikasikan status stok memperoleh dukungan dari hasil pengujian karena ketiga kategori dapat diprediksi pada seluruh skenario. Bagian hipotesis yang menyebut tingkat akurasi baik perlu ditafsirkan secara terbatas karena nilai tertinggi 70,83% masih disertai tujuh kesalahan pada skenario 90:10. Model layak dipandang sebagai alat bantu analitis, bukan sebagai pengganti keputusan manajemen persediaan. Pengembangan

berikutnya dapat menggunakan periode data lebih panjang, atribut operasional tambahan, validasi berulang, serta perbandingan dengan algoritma lain. Seleksi fitur dan rekayasa fitur juga dapat digunakan untuk mencari kombinasi prediktor yang lebih representatif. Sistem pengelolaan stok berbasis digital akan memberi manfaat lebih besar ketika data transaksi, pembelian, dan persediaan terhubung dalam struktur yang konsisten. Pengembangan sistem stok dan penjualan berbasis web telah dipandang mendukung transformasi pengelolaan data operasional (Samputri & Halimah, 2026).

## KESIMPULAN

Klasifikasi status stok berhasil dibangun dari 250 catatan histori penjualan PT Berlian Djaya Nusantara melalui tahapan pemilihan data, pembersihan, transformasi, pembentukan target berbasis kuartil, pembagian data, pemodelan Naive Bayes, dan evaluasi confusion matrix. Nilai Q1 sebesar 138 dan Q3 sebesar 311 menghasilkan 63 data Rendah, 124 data Sedang, dan 63 data Tinggi. Model akhir memakai Jumlah Penjualan, Frekuensi Penjualan, dan Stok Awal sebagai prediktor, sedangkan Stok Akhir hanya digunakan untuk membentuk Status Stok agar tidak terjadi data leakage. Lima rasio pembagian data menghasilkan performa berbeda, dengan hasil terbaik pada 90:10 berupa accuracy 70,83%, precision 70,27%, recall 72,22%, dan F1-score 71,23%. Kategori Sedang paling konsisten dikenali, sedangkan kelas ekstrem masih mengalami kesalahan pada beberapa skenario. Hasil tersebut menjawab kebutuhan klasifikasi dan menunjukkan bahwa histori penjualan dapat diubah menjadi informasi kategori untuk mendukung peninjauan persediaan. Keterbatasan utama berasal dari jumlah data yang hanya 250, penggunaan tiga prediktor, satu kali pembagian pada setiap rasio, serta asumsi independensi antaratribut. Pengembangan berikutnya perlu mempertimbangkan periode lebih panjang, variabel lead time, stok minimum, stok maksimum, jumlah pembelian, pemasok, permintaan musiman, dan reorder point, disertai validasi berulang serta perbandingan algoritma agar performa lebih stabil. Penggunaan evaluasi multimetri tetap diperlukan agar kemampuan model tidak dinilai hanya melalui akurasi (Han et al., 2022).

## REFERENSI

Aditiya, R., Defit, S., & Nurcahyo, G. W. (2020). Prediksi tingkat ketersediaan stock sembako menggunakan algoritma FP-Growth dalam meningkatkan penjualan. Jurnal Informatika Ekonomi Bisnis, 67–73. https://doi.org/10.37034/infeb.v2i3.44

Andriansyah, R. (2021). Penerapan algoritma Naive Bayes untuk klasifikasi penjualan barang pada perusahaan retail. Jurnal Sistem Informasi dan Teknologi Informasi, 9(2), 112–120.

Faisal, et al. (2025). Perkembangan teknologi informasi dan transformasi digital dalam sektor distribusi barang. Jurnal Distribusi dan Logistik Digital, 1(1), 1–10.

Giovanno, & Widyadana. (2022). Diskretisasi berbasis kuartil untuk pembentukan variabel target status stok. Jurnal Optimasi Sistem Industri, 14(1), 12–20.

Han, J., Kamber, M., & Pei, J. (2022). Data mining: Concepts and techniques (4th ed.). Morgan Kaufmann.

Heizer, J., Render, B., & Munson, C. (2022). Operations management: Sustainability and supply chain management (14th ed.). Pearson Education.

Hidayat, M. (2024). Implementasi data mining dalam manajemen persediaan barang pada perusahaan distribusi. Jurnal Teknologi Informasi dan Bisnis Digital, 8(1), 44–55.

Julianto, A., & Andayani, S. (2024). Penerapan data mining menggunakan data historis penjualan untuk pengelolaan persediaan barang. Jurnal Teknologi Informasi dan Sistem Informasi, 1(1), 1–10.

Lubis, B., & Sitohang, S. (2023). Peningkatan kualitas data untuk optimasi model klasifikasi status stok penjualan. Jurnal Ilmu Komputer dan Teknologi Informasi, 15(2), 45–52.

Manajemen, A., & Sulistyawati, E. (2024). Pengelolaan persediaan dan efisiensi operasional pada perusahaan distribusi. Jurnal Manajemen dan Logistik, 8(1), 45–58.

Pahlevi, M. R. (2026). Out of stock (OOS) management strategies to improve warehouse service levels at PT XYZ Malang Branch. Jurnal Sekretaris & Administrasi Bisnis (JSAB), 10(1), 1–12. https://doi.org/10.31104/jsab.v10i1.507

Pratama, D., & Nugroho, A. (2022). Optimalisasi stok barang menggunakan pendekatan data mining pada perusahaan distribusi. Jurnal Informatika Bisnis, 14(2), 67–78.

Ramadhan, F., & Putri, S. (2023). Prediksi kebutuhan stok barang menggunakan metode klasifikasi pada perusahaan distribusi. Jurnal Sistem Cerdas dan Analitika Data, 5(1), 23–35.

Saepudin, S., Widiastuti, S., & Irawan, C. (2023). Sentiment analysis of social media platform reviews using the Naive Bayes classifier algorithm. Jurnal Sisfokom (Sistem Informasi dan Komputer), 12(2), 236–243. https://doi.org/10.32736/sisfokom.v12i2.1650

Samputri, D., & Halimah. (2026). Pengembangan sistem manajemen stok dan penjualan berbasis web dalam mendukung transformasi digital menggunakan metode prototype pada Butik Merry's Fashion Lampung. Jurnal Algoritma, 23(1). https://doi.org/10.33364/algoritma/v.23-1.3432

Sari, S. F., & Nasution, Y. R. (2025). Optimalisasi manajemen stok barang menggunakan metode Apriori berbasis data mining. Jurnal Pendidikan dan Teknologi Indonesia, 5(1), 215–226. https://doi.org/10.52436/1.jpti.631

Wijaya, A. (2023). Implementasi algoritma Naive Bayes pada sistem persediaan barang berbasis data historis penjualan. Jurnal Komputer dan Informatika Indonesia, 7(3), 145–156.