

## Model Prediksi Jumlah Penjualan Pelumas Mesin Di PT. X Dengan Algoritma *Naïve Bayes*

 Nilam Purnama<sup>1</sup>, Fitri Insani<sup>\*2</sup>, Elin Haerani<sup>3</sup>, Iis Afrianty<sup>4</sup>

 Email: [1nilampurnama1503@gmail.com](mailto:nilampurnama1503@gmail.com), [2fitri.insani@uin-suska.ac.id](mailto:fitri.insani@uin-suska.ac.id), [3elin.haerani@uin-suska.ac.id](mailto:elin.haerani@uin-suska.ac.id), [4iis.afrianty@uin-suska.ac.id](mailto:iis.afrianty@uin-suska.ac.id)
<sup>1,2,3,4</sup>Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau

Diterima: 30 November 2024 | Direvisi: - | Disetujui: 20 Desember 2024

 ©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,  
Universitas Muhammadiyah Riau, Indonesia

### Abstrak

Pelumas mesin merupakan bahan penting untuk mengurangi gesekan antara dua permukaan yang bergerak, meningkatkan efisiensi kerja mesin, dan memperpanjang usia komponen. Penelitian ini bertujuan untuk memprediksi jumlah penjualan pelumas mesin di PT. X menggunakan algoritma *Naïve Bayes*. Data yang digunakan meliputi atribut seperti tahun, bulan, deskripsi material, total alokasi, realisasi, dan sisa alokasi, dengan jumlah data sebanyak 3.006 yang diperoleh dari sistem *Warehouse Management System* (WMS) PT. X. Model diuji dengan metode *10-Fold Cross Validation* dan pengujian tanpa validasi tersebut. Hasil pengujian menunjukkan akurasi sebesar 71% dengan *10-Fold Cross Validation*, dibandingkan dengan 14% tanpa validasi. Pengujian tambahan menunjukkan akurasi sebesar 5%, dengan RMSE 124,71 dan MAPE 0,95. Berdasarkan hasil penelitian ini, disarankan untuk melakukan optimisasi preprocessing data, seperti penanganan ketidakseimbangan data dan normalisasi fitur, agar akurasi prediksi dapat meningkat. Selain itu, penggunaan teknik validasi yang lebih beragam, seperti stratified cross-validation, dapat memberikan evaluasi yang lebih stabil. Mengingat bahwa prediksi hanya dipengaruhi oleh data historis, disarankan untuk memperbarui data secara berkala agar model tetap relevan dan akurat. Penelitian ini diharapkan dapat membantu PT. X dalam merencanakan strategi penjualan dan pengelolaan stok pelumas secara lebih efektif.

**Kata kunci:** pelumas mesin, *naïve bayes*, *confusion matrix*, *10-fold cross validation*, *warehouse management system*

### *Prediction Model for the Number of Machine Lubricant Sales at PT. X With the Naïve Bayes Algorithm*

#### Abstract

*Machine lubricants are essential materials used to reduce friction between two moving surfaces, improve machine efficiency, and extend the lifespan of components. This study aims to predict the sales volume of machine lubricants at PT. X using the Naïve Bayes algorithm. The data used includes attributes such as year, month, material description, total allocation, realization, and remaining allocation, with a total of 3,006 data points obtained from PT. X's Warehouse Management System (WMS). The model was tested using the 10-Fold Cross Validation method and testing without such validation. The test results show an accuracy of 71% with 10-Fold Cross Validation, compared to 14% without validation. Additional testing showed an accuracy of 5%, with RMSE of 124.71 and MAPE of 0.95. Based on these results, it is recommended to optimize data preprocessing, such as handling data imbalance and feature normalization, to improve prediction accuracy. Furthermore, using more diverse validation techniques, such as stratified cross-validation, can provide more stable evaluations. Given that predictions are influenced solely by historical data, it is recommended to periodically update the data to keep the model relevant and accurate. This research is expected to assist PT. X in planning sales strategies and managing lubricant stock more effectively.*

**Keywords:** *machine lubricant, naïve bayes, confusion matrix, 10-fold cross validation, warehouse management system*

## 1. PENDAHULUAN

Pelumas mesin adalah bahan yang digunakan untuk mengurangi gesekan antara dua permukaan yang bergerak. Fungsinya adalah untuk memastikan mesin tetap bekerja dengan baik dan memperpanjang usia komponen mekanis. Dalam industri, pelumas mesin

sangat penting untuk menjaga agar mesin beroperasi dengan lancar dan efisien[1]. Penjualan pelumas mesin menjadi aspek vital bagi perusahaan distribusi seperti PT. X, yang harus mampu memprediksi permintaan pasar untuk mengoptimalkan strategi penjualan dan pengelolaan stok[2]. Meskipun banyak faktor yang dapat mempengaruhi permintaan, seperti fluktuasi harga dan variasi produk, penelitian ini berfokus pada pengembangan model prediksi yang didasarkan pada data historis penjualan[3].

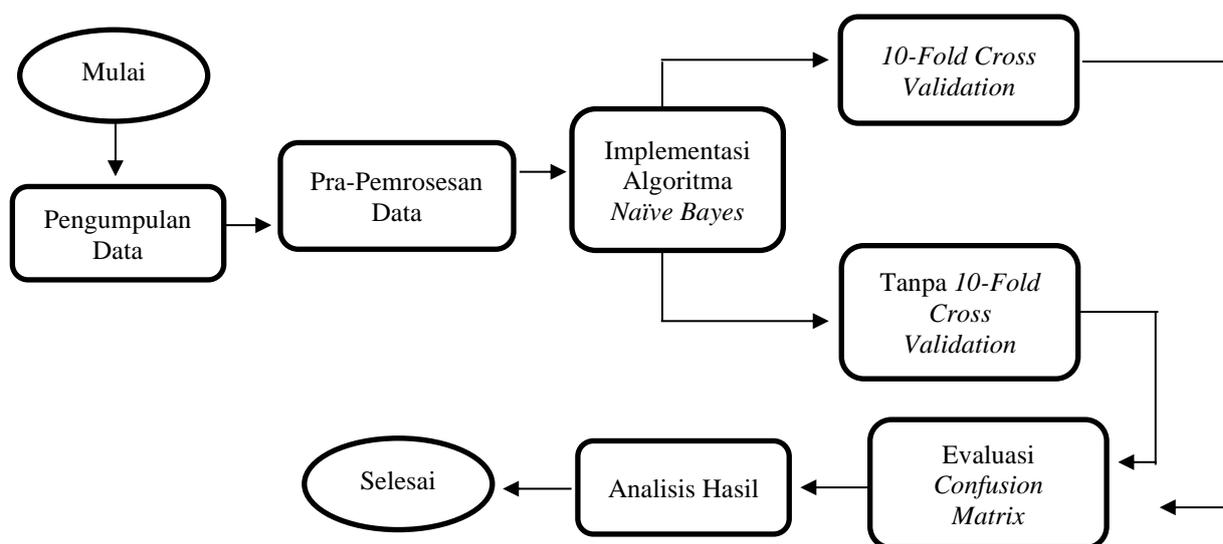
Beberapa penelitian telah dilakukan untuk memprediksi penjualan menggunakan algoritma machine learning seperti *Naïve Bayes*, *Decision Tree*, dan *SVM*. Berbagai penelitian menunjukkan bahwa algoritma pembelajaran mesin, termasuk *Naïve Bayes*, dapat digunakan untuk memprediksi penjualan di berbagai sektor industri[4]. Contohnya, penelitian oleh Made Adrian Astalina Pramana yang berjudul “Penerapan Algoritma *Naïve Bayes* Untuk Prediksi Penjualan Produk Terlaris Pada CV Akusara Jaya Abadi” memperoleh tingkat akurasi sebesar 89,33% menunjukkan bahwa *Naïve Bayes* efektif dalam menangani dataset dengan banyak variabel independen serta mampu memberikan hasil yang cepat dan akurat[5][6][7]. Selain itu, teknik evaluasi seperti *Confusion Matrix* terbukti membantu dalam menilai keakuratan model prediksi. Namun, penerapan algoritma ini dalam konteks penjualan pelumas mesin masih terbatas, sehingga memberikan peluang untuk eksplorasi lebih lanjut dalam pengembangan model prediksi yang lebih efektif dan spesifik[8].

Algoritma *Naïve Bayes* dipilih dalam penelitian ini karena kesederhanaannya, kecepatan dalam memproses data besar, dan kemampuannya menangani data dengan banyak variabel independen[6]. Algoritma ini dikenal efektif dalam menghasilkan prediksi yang cepat dan akurat. Namun, penerapan *Naïve Bayes* dalam konteks penjualan pelumas mesin masih jarang dilakukan, sehingga membuka peluang untuk eksplorasi lebih lanjut. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model prediksi yang dapat membantu PT. X dalam merencanakan strategi penjualan dan pengelolaan stok pelumas mesin secara lebih efektif[9]. Dengan memanfaatkan *10-Fold Cross Validation*, diharapkan model ini mampu memberikan prediksi yang lebih andal dan akurat dalam mengidentifikasi pola permintaan berdasarkan data historis.[10].

Dengan latar belakang tersebut, penelitian ini bertujuan untuk menjawab beberapa pertanyaan. Pertama, bagaimana algoritma *Naïve Bayes* dapat diterapkan untuk memprediksi penjualan pelumas mesin di PT. X? Kedua, seberapa akurat model prediksi yang dihasilkan dalam memprediksi jumlah penjualan pelumas mesin? Penilaian model dilakukan dengan dua metode, yaitu menggunakan *10-Fold Cross Validation* dan tanpa validasi tersebut, untuk memastikan bahwa model dapat diandalkan dalam pengambilan keputusan. Hasil penelitian ini diharapkan dapat memberikan kontribusi bagi pengembangan metode prediksi penjualan yang lebih efektif dan efisien di sektor industri pelumas mesin, sehingga membantu PT. X dalam mengelola stok dan merencanakan strategi penjualan secara optimal[11].

## 2. METODE PENELITIAN

Metodologi penelitian ini dilakukan dengan pendekatan kuantitatif menggunakan algoritma *Naïve Bayes* untuk memprediksi penjualan pelumas mesin di PT. X. Penelitian ini tidak hanya berfokus pada pembuatan model, tetapi juga pada evaluasi model menggunakan teknik pengujian seperti *Confusion Matrix* dan *10-Fold Cross Validation*[12]. Berikut adalah *flowchart* tahapan penelitian yang ditunjukkan pada Gambar 1.



Gambar 1 Flowchart Tahap Penelitian

Langkah-langkah yang dilakukan di metodologi penelitian pada Gambar 1 dapat dijelaskan sebagai berikut:

### 2.1. Pengumpulan Data

Tahap awal adalah mengumpulkan data historis penjualan pelumas mesin dari PT. X. Proses pengumpulan data ini melibatkan ekstraksi informasi langsung dari aplikasi *Warehouse Management System (WMS)* yang didapat dari PT X, yang mencatat berbagai informasi penjualan pelumas mesin. Data ini merupakan pondasi untuk analisis yang lebih lanjut[13].

### 2.2. Pra-Pemrosesan Data

Setelah data berhasil dikumpulkan, langkah berikutnya adalah melakukan pra-pemrosesan. Tahapan pra-pemrosesan meliputi:

1. Pembersihan Data: Menghapus entri yang tidak valid, seperti data yang memiliki nilai hilang atau duplikat[14].
2. Transformasi Data: Mengubah format data menjadi numerik jika diperlukan, dan melakukan normalisasi untuk memastikan semua fitur berada dalam ukuran yang konsisten[15].
3. *Encoding* Variabel Kategorikal: Mengubah variabel kategorikal menjadi format numerik agar dapat digunakan oleh algoritma *Naïve Bayes*[16].

### 2.3. Pembagian Data

Data yang telah diproses dibagi menjadi dua bagian yaitu data latih dan data uji. Pada pembagian data ini dilakukan dengan menggunakan 2 skema. Yang pertama dengan menggunakan *10-Fold Cross Validation*, yang kedua dengan menggunakan perbandingan 70:30, 70% data latih, 30% sisanya akan digunakan untuk menguji seberapa akurat model yang telah dibuat[17]. Pengujian model menggunakan *10-Fold Cross Validation* dipilih karena metode ini memberikan evaluasi yang lebih akurat dan andal dibandingkan dengan pembagian data secara acak satu kali. Dengan membagi dataset menjadi sepuluh lipatan (folds), setiap lipatan secara bergantian digunakan sebagai data uji, sementara sisanya digunakan untuk melatih model. Pendekatan ini memastikan bahwa setiap data dalam dataset digunakan untuk pengujian dan pelatihan, sehingga meningkatkan keakuratan hasil evaluasi. Selain itu, metode ini membantu mengurangi variansi hasil dan mencegah overfitting, di mana model terlalu menyesuaikan diri pada data latih tetapi kurang efektif pada data baru. Dalam penelitian ini, *10-Fold Cross Validation* menghasilkan akurasi yang jauh lebih tinggi, yaitu 71%, dibandingkan hanya 14% tanpa *10-Fold Cross Validation*, menunjukkan urgensi penggunaan metode ini untuk memastikan keandalan model prediksi penjualan pelumas mesin di PT. X.

### 2.4. Implementasi Algoritma *Naïve Bayes*

Implementasi algoritma *Naïve Bayes* dalam penelitian ini bertujuan untuk memprediksi jumlah penjualan pelumas mesin di PT. X berdasarkan data historis yang telah diproses[18]. *Naïve Bayes* adalah algoritma klasifikasi yang didasarkan pada Teorema Bayes, dengan asumsi bahwa setiap fitur saling independen. Algoritma ini banyak digunakan dalam klasifikasi karena kesederhanaannya dan efektivitasnya, terutama saat menangani dataset besar[18]. Data dibagi menjadi 80% untuk data latih dan 20% untuk data uji.

Cara kerjanya mengikuti *Teorema Bayes*, yang menyatakan bahwa probabilitas suatu kejadian dapat dihitung berdasarkan informasi yang ada. Rumus dasar dari *Naïve Bayes* adalah sebagai berikut:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \tag{1}$$

Di mana  $P(C|X)$  adalah probabilitas kelas C diberikan fitur X,  $P(X|C)$  adalah probabilitas fitur X diberikan kelas C,  $P(C)$  adalah probabilitas kelas C,  $P(X)$  adalah probabilitas fitur X.

### 2.5. Pengujian

Setelah model dibangun, pengujian dilakukan menggunakan dua skema yaitu dengan *10-Fold Cross Validation* dan yang tidak. Dua skema itu akan di evaluasi dengan menggunakan *Confusion Matrix*, *Confusion Matrix* digunakan untuk mengukur seberapa akurat model dalam memprediksi hasil penjualan yang tepat[19]. Matriks ini menggambarkan perbandingan antara hasil prediksi dan hasil sebenarnya dari data uji. Dengan *Confusion Matrix*, kita dapat mengetahui nilai *true positive*, *true negative*, *false positive*, dan *false negative*, yang kemudian digunakan untuk menghitung metrik akurasi, presisi, dan recall[20]. Berikut adalah tabel *Confusion Matrix* yang ditunjukkan pada Tabel 1.

Tabel 1 Tabel *Confusion Matrix*

	Prediksi Naik	Prediksi Turun
Sebenarnya Naik	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
Sebenarnya Turun	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} * 100 \tag{2}$$

$$\text{Presisi} = \frac{TP}{TP + FP} * 100 \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} * 100 \tag{4}$$

$$F-I \text{ Score} = \frac{\text{Recall} * \text{Presisi}}{\text{Recall} + \text{Presisi}} * 2 \tag{5}$$

### 2.6. Analisis Hasil

Setelah melakukan pengujian dan evaluasi, hasil analisis akan disajikan untuk mengevaluasi performa model *Naïve Bayes*. Hasil akurasi dari kedua metode pengujian (*10-Fold Cross Validation* dan yang tidak) dibandingkan untuk menentukan efektivitas model dalam memprediksi penjualan pelumas mesin[21]. Analisis ini juga mencakup diskusi mengenai faktor-faktor yang mempengaruhi hasil, serta rekomendasi untuk perbaikan model di masa depan[22].

Dengan mengikuti metodologi ini, diharapkan model prediksi yang dihasilkan dapat memberikan kontribusi nyata dalam perencanaan strategi penjualan dan pengelolaan stok pelumas mesin di PT. X.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Pengumpulan Data

Pengumpulan data pada penelitian ini diambil dari penjualan pelumas mesin di PT. X antara Januari 2023 hingga Januari 2024, dengan total 3.006 transaksi. Proses pengumpulan data ini melibatkan ekstraksi informasi langsung dari aplikasi *Warehouse Management System (WMS)* yang didapat dari PT X, yang mencatat berbagai informasi penjualan pelumas mesin. Data ini merupakan data mentah yang belum diproses. Data yang dikumpulkan ditunjukkan pada Tabel 2.

Tabel 2 Tabel Penjualan Pelumas mesin

NO	Marketing - Allocation			Alokasi Awal	Alokasi Tambahan	Total Alokasi		Realisasi		Sisa Alokasi	
	Year	Month	Material Description			Qty	Conv. Liter	Qty	Conv. Liter	Qty	Conv. Liter
1	2023	Januari	FASTRON	116	0	116	2784	116	2784	0	0
2	2023	Januari	ENDURO	10	0	10	60	10	60	0	0
3	2023	Februari	RORED	25	0	25	600	0	0	25	600
4	2023	Februari	MEDITRAN	6741	0	6741	134820	4771	95420	1970	39400
5	2023	Maret	FASTRON	4	0	4	96	4	96	0	0
6	2023	Maret	MEDITRAN	4209	0	4209	84180	2481	49620	1728	34560
...	...	...	...	...	...	...	...	...	...	...	...
3.006	2024	Februari	FASTRON	4	0	4	96	4	96	...	0

### 3.2. Pra-Pemrosesan Data

Dalam tahap pra-pemrosesan, ada tiga langkah yang dilakukan, yaitu pembersihan data, transformasi data, dan *Encoding variable* kategorikal. Pada tahap transformation, data numerik seperti Total Alokasi Conv. Liter, Realisasi Conv. Liter, dan Sisa Alokasi Conv. Liter diolah untuk digunakan dalam analisis[23].

#### a. Pembersihan Data

Pada langkah pembersihan, data yang tidak relevan dan atribut yang tidak diperlukan dihapus, sehingga hanya menyisakan atribut *Year*, *Month*, *Material Description*, *Total Alokasi Conv.Liter*, *Realisasi Conv. Liter*, dan *Sisa Alokasi Conv.Liter*[23]. Berikut adalah hasil setelah melakukan tahapan data *Cleaning*, yang ditunjukkan dalam Tabel 3.

Tabel 3 Hasil Setelah *Cleaning* Data

NO	Marketing - Allocation			Total Alokasi	Realisasi	Sisa Alokasi
	Year	Month	Material Description	Conv.Liter	Conv.Liter	Conv.Liter
1	2023	Januari	FASTRON	2784	2784	0
2	2023	Februari	MEDITRAN	134820	95420	39400
3	2023	Maret	MEDITRAN	84180	49620	34560
...	...	...	...	...	...	...
1.646	2024	Januari	FASTRON	96	96	0

Berdasarkan Tabel 3, atribut *Year* dan *Month* menunjukkan periode transaksi, *Material Description* adalah jenis produk pelumas mesin yang dijual, *Total Alokasi Conv.Liter* merupakan jumlah yang dialokasikan, *Realisasi Conv. Liter* menunjukkan jumlah penjualan yang sebenarnya, dan *Sisa Alokasi Conv.Liter* merupakan sisa penjualan yang dialokasikan[24].

b. Transformasi Data

Selanjutnya, pada tahap transformasi data, dilakukan beberapa operasi untuk membuat data lebih siap untuk analisis, yaitu mengubah format bulan dari teks menjadi angka untuk memudahkan analisis lebih lanjut. Berikut adalah hasil setelah transformasi data yang ditunjukkan pada Tabel 4[25].

Tabel 4 Hasil Setelah Transformasi Data

NO	Marketing - Allocation			Total Alokasi	Realisasi	Sisa Alokasi
	Year	Month	Material Description	Conv.Liter	Conv.Liter	Conv.Liter
1	2023	1	FASTRON	2784	2784	0
2	2023	2	MEDITRAN	134820	95420	39400
3	2023	3	MEDITRAN	84180	49620	34560
...	...	...	...	...	...	...
1.646	2024	1	FASTRON	96	96	0

c. Encoding Variabel Kategorikal

Pada langkah ini, menggunakan *One-Hot Encoding* untuk mengonversi kolom kategorikal seperti "*Material Description*" menjadi format numerik yang dapat digunakan oleh algoritma pembelajaran mesin. Dengan teknik ini, setiap kategori diwakili oleh kolom biner, memungkinkan model untuk memahami data kategorikal dengan lebih baik[26]. Setelah semua tahapan ini selesai, dataset siap untuk dianalisis lebih lanjut atau digunakan dalam pembangunan model prediktif. Berikut adalah codingan untuk *One-Hot Encoding* yang ditunjukkan pada Gambar 2.

```

('cat', OneHotEncoder(drop='first'), ['Material Description']) # Mengonversi 'Material
Description' ke numerik
    ]
    
```

Gambar 2 Implementasi One-Hot Encoding

3.3. Menganalisa Data Menggunakan *Naïve Bayes*

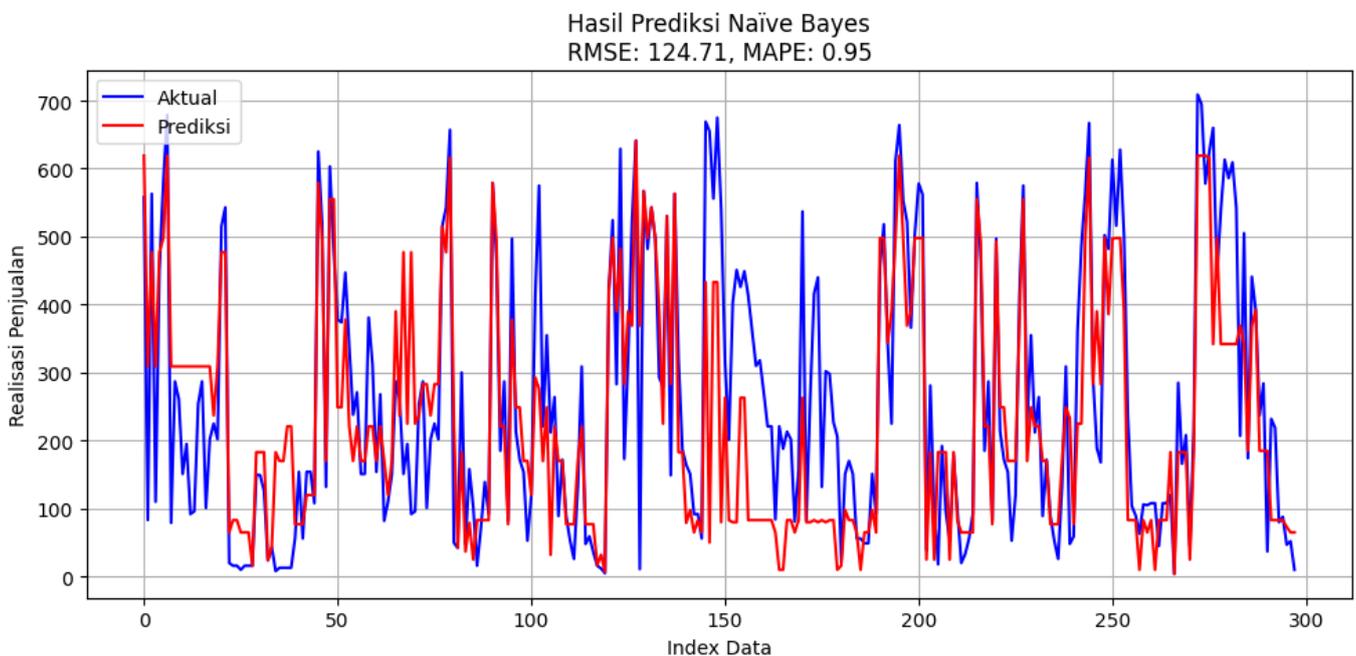
Setelah tahap pra-pemrosesan data, analisis dilakukan dengan menerapkan algoritma *Naïve Bayes* menggunakan *Google Colab*. Algoritma *Naïve Bayes* diterapkan untuk memprediksi realisasi penjualan pelumas menggunakan dataset yang telah diproses sebelumnya. Model ini dipilih karena kemampuannya menangani data berukuran besar dengan efisien. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian, guna memastikan model dapat diuji secara valid.

Hasil evaluasi menunjukkan bahwa model memiliki akurasi sebesar 5%, RMSE sebesar 124.71, dan MAPE sebesar 0.95%. Akurasi yang rendah menunjukkan bahwa model menghadapi kesulitan dalam memprediksi dengan tepat nilai realisasi penjualan. Namun, nilai MAPE yang relatif rendah menunjukkan bahwa rata-rata kesalahan persentase prediksi cukup kecil.

Tabel 5 Hasil Prediksi *Naive Bayes*

Tahun	Bulan	Jenis Pelumas	Prediksi Realisasi	Realisasi Aktual	Status
2023	8	Meditran	619	558	Meningkat
2023	8	Meditran	309	83	Meningkat
2023	8	Meditran	477	563	Menurun
2023	8	Meditran	309	110	Meningkat
2023	8	Meditran	477	452	Meningkat

Pada Tabel 5 hasil prediksi menunjukkan bahwa ada status meningkat dan menurun, apabila prediksi lebih tinggi dari realisasi aktual, maka statusnya meningkat. Namun, apabila realisasi aktual lebih tinggi, maka statusnya menurun. Hal ini menunjukkan bahwa model masih memiliki kesalahan yang signifikan dalam prediksi data kontinu seperti penjualan pelumas.



Gambar 3 Hasil Prediksi *Naive Bayes*

Gambar 3 menunjukkan perbandingan antara prediksi realisasi penjualan dengan realisasi aktual menggunakan algoritma *Naive Bayes*. Pada grafik ini terlihat bahwa meskipun terdapat beberapa kesamaan pola antara prediksi dan aktual, perbedaan yang signifikan muncul pada beberapa titik data. Hal ini mengindikasikan bahwa model belum sepenuhnya mampu menangkap pola kompleks dalam data penjualan. Kesalahan ini diperkuat dengan nilai RMSE sebesar 124.71 dan MAPE sebesar 0.95%, yang menunjukkan bahwa prediksi model masih jauh dari hasil aktual. Visualisasi ini memberikan gambaran jelas tentang kelemahan *Naive Bayes* dalam menangani data kontinu seperti penjualan pelumas mesin.

Berdasarkan hasil analisis, algoritma *Naive Bayes* memberikan prediksi yang cukup terstruktur, meskipun akurasi masih rendah. Hal ini kemungkinan disebabkan oleh karakteristik data kontinu yang kurang sesuai dengan asumsi independensi fitur pada *Naive Bayes*. Meskipun demikian, algoritma ini tetap memberikan gambaran tren yang berguna, seperti status meningkat atau menurun. Selain mengimplementasikan algoritma *Naive Bayes*, dapat juga diterapkan *10-Fold Cross Validation* untuk memperoleh hasil evaluasi yang lebih akurat dan mengurangi bias pada hasil[27].

BEGIN

```
INISIALISASI df <- BacaDataset('semhas.csv')
```

```
INISIALISASI preprocessor <- ColumnTransformer DENGAN:
```

- StandardScaler UNTUK kolom numerik
- OneHotEncoder UNTUK 'Material Description'

INISIALISASI kf <- KFold(n\_splits=10, shuffle=TRUE, random\_state=42)

INISIALISASI model\_pipeline <- BuatPipeline DENGAN:

- preprocessor
- GaussianProcessRegressor

PREDIKSI y\_pred\_class <- model\_pipeline PREDIKSI data\_uji

HITUNG conf\_matrix <- confusion\_matrix(y\_class, y\_pred\_class)

HITUNG accuracy <- accuracy\_score(y\_class, y\_pred\_class)

HITUNG precision <- precision\_score(y\_class, y\_pred\_class, pos\_label="Meningkat")

HITUNG recall <- recall\_score(y\_class, y\_pred\_class, pos\_label="Meningkat")

TAMPILKAN Heatmap(conf\_matrix) DENGAN label dan judul

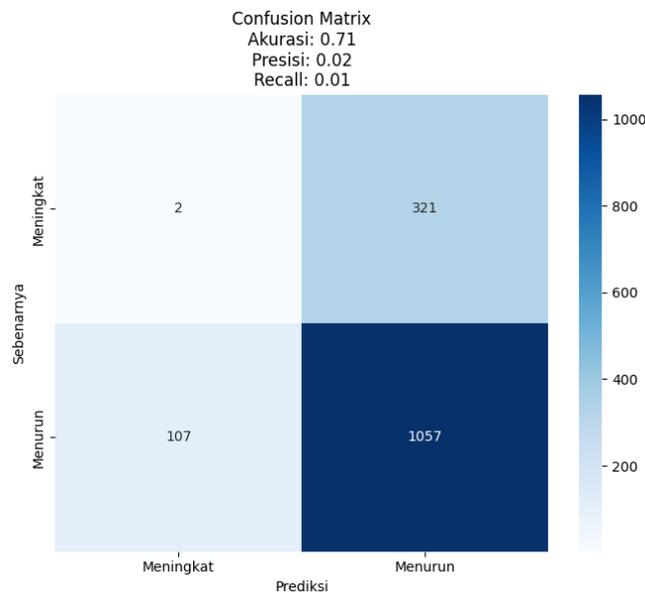
TAMPILKAN grafik

CETAK "Akurasi:", accuracy

CETAK "Presisi (Meningkat):", precision

CETAK "Recall (Meningkat):", recall

END



Gambar 4 Hasil dari 10-Fold Cross Validation

Dengan hasil dari confusion matrix pengujian dengan 10-Fold Cross Validation, maka didapatkan nilai akurasi sejumlah :

$$\text{Akurasi} = \frac{2+1057}{2+157+321+107} * 100 = 71,2\%=71\%$$

Gambar 4 menunjukkan hasil dari 10-Fold Cross Validation yang menunjukkan akurasi sebesar 71%.

Yang kedua, menggunakan model uji dengan teknik split validation 70:30, 70% data latih dan 30% data uji.

BEGIN

INISIALISASI df <- BacaDataset('semhas.csv')

INISIALISASI preprocessor <- ColumnTransformer DENGAN:

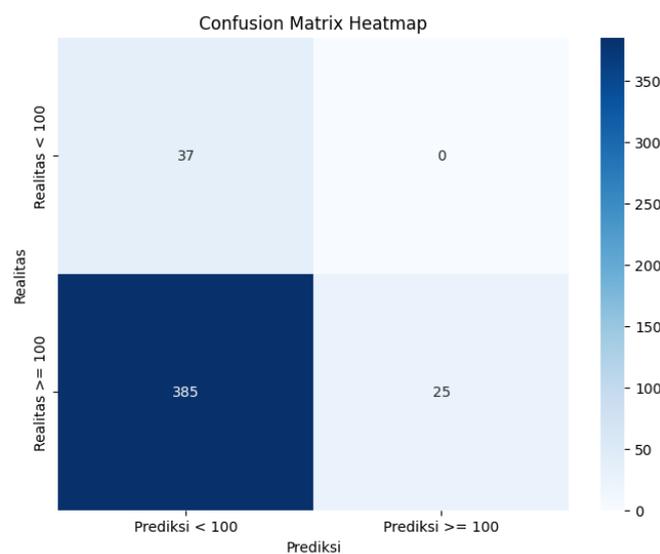
- StandardScaler UNTUK kolom numerik
- OneHotEncoder UNTUK 'Material Description'

INISIALISASI model\_pipeline <- BuatPipeline DENGAN:

- preprocessor
- GaussianProcessRegressor

```
BAGI data MENJADI X_train, X_test, y_train, y_test DENGAN train_test_split(X, y, test_size=0.1)
FIT model_pipeline PADA data Latih (X_train, y_train)
PREDIKSI y_pred <- model_pipeline PREDIKSI data uji (X_test)
TAMPILKAN hasil_prediksi <- Salin X_test DAN TAMBAHKAN 'Prediksi' DARI y_pred
CETAK hasil_prediksi
KLASIFIKASI y_test_bin <- Konversi y_test KE biner DENGAN nilai ambang 100
KLASIFIKASI y_pred_bin <- Konversi y_pred KE biner DENGAN nilai ambang 100
HITUNG cm <- confusion_matrix(y_test_bin, y_pred_bin)
TAMPILKAN Heatmap(cm)
TAMPILKAN grafik
CETAK classification_report(y_test_bin, y_pred_bin)
```

END



Gambar 5 Hasil dari Teknik *split validation* 70:30

Dengan hasil dari teknik *split validation* 70:30 yang ditunjukkan pada Gambar 5 maka didapatkan nilai akurasi sejumlah :

$$\text{Akurasi} = \frac{37 + 25}{37 + 25 + 0 + 385} * 100 = 13,8\% = 14\%$$

### 3.4. Perbandingan Hasil Akurasi

Setelah semua langkah pengujian dilakukan, selanjutnya dilakukan perbandingan hasil analisis dari algoritma *Naïve Bayes* dengan dua metode pengujian yang berbeda. Hasil perbandingan ini akan disajikan dalam bentuk tabel[20]. Yang akan dibandingkan adalah nilai akurasi dari pengujian menggunakan *10-Fold Cross Validation* dan yang tidak menggunakan *10-Fold Cross Validation*[28]. Hasil akurasi dari kedua metode tersebut dapat dilihat pada Tabel 6.

Tabel 6 Hasil Perbandingan Akurasi

No	Metode Pengujian	Akurasi
1	<i>10-Fold Cross Validation</i>	71%
2	Tanpa <i>10-Fold Cross Validation</i> 70:30	14%

Berdasarkan Tabel 6, hasil akurasi menunjukkan bahwa pengujian dengan metode *10-Fold Cross Validation* menghasilkan performa yang lebih baik dibandingkan dengan pengujian tanpa metode tersebut[28]. Hal ini menunjukkan bahwa algoritma *Naïve Bayes* dapat digunakan secara efektif dalam memprediksi jumlah penjualan pelumas mesin di PT. X[19].

#### 4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma *Naïve Bayes* memiliki performa yang cukup baik dalam tugas prediksi, meskipun hasil akurasi yang diperoleh hanya sebesar 5%. Algoritma ini cocok digunakan untuk tugas prediksi dengan data yang sederhana, karena kemudahannya dalam implementasi dan efisiensi prosesnya. Berdasarkan hasil penelitian ini, disarankan untuk melakukan optimalisasi preprocessing data, seperti penanganan ketidakseimbangan data dan normalisasi fitur, agar akurasi prediksi dapat meningkat. Selain itu, penggunaan teknik validasi yang lebih beragam, seperti stratified cross-validation, dapat memberikan evaluasi yang lebih stabil. Mengingat bahwa prediksi hanya dipengaruhi oleh data historis, disarankan juga untuk terus memperbarui data yang digunakan agar model tetap relevan dan akurat dalam memprediksi penjualan. Dengan langkah-langkah ini, diharapkan dapat meningkatkan efektivitas model dalam mendukung pengambilan keputusan di PT. X.

#### DAFTAR PUSTAKA

- [1] A. Hendrawan, A. Sasongko D, A. Maritim, N. Cilacap, and T. Maritim, "PENGARUH UMUR PELUMASAN TERHADAP SUHU MESIN INDUK KM. LOGISTIK NUSANTARA 4," 2021.
- [2] Y. Sulindra and E. Herjanto, "STRATEGI PENJUALAN PRODUK PELUMAS PT PPA DENGAN MENGGUNAKAN ANALISIS SWOT."
- [3] R. Jauhari Aisawa Krismanegara, H. Purnomo, U. Nusantara PGRI, J. K. Achmad Dahlan, and K. Jawa Timur, "Seminar Nasional Manajemen, Ekonomi dan Akuntansi Fakultas Ekonomi dan Bisnis UNP Kediri MENINGKATKAN OMZET PENJUALAN MELALUI OPTIMALISASI KUALITAS PRODUK , VARIASI PRODUK, DAN PERSEPSI HARGA PADA NINDY DEKORASI."
- [4] H. Herman, "*Naïve Bayes* Classifier dan K-Nearest Neighbor Pada Analisis Sentimen Perkuliahan Daring di Universitas Muslim Indonesia," 2024, doi: 10.33096/busiti.v5i1.2202.
- [5] P. Sofyan Zakaria, R. Julianto, and R. Surya Bernada, "IMPLEMENTASI *NAÏVE BAYES* MENGGUNAKAN PYTHON DALAM KLASIFIKASI DATA." [Online]. Available: <https://jurnalmahasiswa.com/index.php/biikma>
- [6] "6498-Article Text-18655-1-10-20231214-2".
- [7] Chely Aulia Misrun, E. Haerani, M. Fikry, and E. Budianita, "Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode *Naïve Bayes* classifier," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 207–215, Apr. 2023, doi: 10.37859/coscitech.v4i1.4790.
- [8] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, "Teknik K-Fold *Cross Validation* untuk Mengevaluasi Kinerja Mahasiswa," *Jurnal Algoritma*, vol. 21, no. 1, May 2024, doi: 10.33364/algoritma/v.21-1.1618.
- [9] A. Supriyadi Sunge and A. Turmudi Zy, "ANALISIS PREDIKSI PENJUALAN DENGAN METODE REGRESI LINEAR DI PT. EAGLE INDUSTRY INDONESIA," 2023.
- [10] S. M. Hudzaifah *et al.*, "Implementasi Algoritma *Naïve Bayes* dalam Memprediksi Tingkat Kelulusan Siswa pada Sertifikasi Mikrotik Certified Network Associate (MTCNA)," 2024. [Online]. Available: <https://journal.stmiki.ac.id>
- [11] M. Safii, B. Efendi Damanik, and G. Artikel, "Algoritma *Naïve Bayes* Untuk Memprediksi Penjualan Pada Toko VJCakes Pematang Siantar *Naïve Bayes* Algorithm For Predicting Sales at the Pematang Siantar VJCakes Store Article Info ABSTRAK," *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, vol. 1, no. 4, pp. 2828–9099, 2022, doi: 10.55123/jomlai.v1i4.1674.
- [12] Rovidatul, Y. Yunus, and G. W. Nurcahyo, "Perbandingan algoritma c4.5 dan *Naïve Bayes* dalam prediksi kelulusan mahasiswa," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 193–199, Apr. 2023, doi: 10.37859/coscitech.v4i1.4755.
- [13] I. G. A. P. A. Putri and I. N. Nurcaya, "PENERAPAN *WAREHOUSE MANAGEMENT SYSTEM* PADA PT UNIPLASTINDO INTERBUANA BALI," *E-Jurnal Manajemen Universitas Udayana*, vol. 8, no. 12, p. 7216, Dec. 2019, doi: 10.24843/ejmunud.2019.v08.i12.p16.
- [14] S. Anwar and H. Lubis, "MENINGKATKAN KUALITAS BIG DATA MELALUI REKAYASA DAN PEMBERSIHAN DATA YANG EFEKTIF." [Online]. Available: <https://www.researchgate.net/publication/376720216>
- [15] V. T. Heidiyanti, "EVALUASI SISTEM MANAJEMEN MUTU DALAM TRANSFORMASI." [Online]. Available: <https://www.researchgate.net/publication/376619056>
- [16] T. Z. Maulani, Z. K. Simbolon, and A. Amirullah, "Implementasi Algoritma *Naïve Bayes* Classifier Dalam Menentukan Topik Tugas Akhir Mahasiswa Berbasis Web," *Jurnal Infomedia*, vol. 4, no. 1, p. 33, Sep. 2019, doi: 10.30811/jim.v4i1.1107.
- [17] W. Abdun Naseer and B. Budi Wahono, "GRADIENT BOOSTING OPTIMIZATION WITH PRUNING TECHNIQUE FOR PREDICTION OF BMT AL-HIKMAH PERMATA CUSTOMER DATA," *Jurnal Informatika Teknologi dan Sains*.

- [18] N. Aini *et al.*, “Implementasi Algoritma *Naïve Bayes* untuk Memprediksi Penjualan Lampu Pada Toko Satria,” *INNOVATIVE: Journal Of Social Science Research*, vol. 3, pp. 9373–9387, 2023.
- [19] I. Widhi Saputro and B. Wulan Sari, “Uji Performa Algoritma *Naïve Bayes* untuk Prediksi Masa Studi Mahasiswa *Naïve Bayes* Algorithm Performance Test for Student Study Prediction,” *Citec Journal*, vol. 6, no. 1, 2019.
- [20] D. Arisandi and T. Sutrisno, “EVALUASI PENILAIAN KINERJA KARYAWAN DENGAN METODE NAÏVE BAYES,” 2024.
- [21] T. Rivanie, R. Pebrianto, T. Hidayat, A. Bayhaqy, W. Gata, and B. Novitasari, “ANALISIS SENTIMEN TERHADAP KINERJA MENTERI KESEHATAN INDONESIA SELAMA PANDEMI COVID-19,” 2021.
- [22] F. Hayatun Nufus, G. Santoso, A. Al Bahij, and U. Muhammadiyah Jakarta, “Pengembangan Karakter Kreatif Dalam Menghadapi Tantangan Abad 21,” 2023.
- [23] A. Azkiya, O. Rohaeni, F. Badruzzaman, and E. Harahap, “Analisis Peramalan Penjualan Produk X Pada Perusahaan XYZ Menggunakan Metode Double Exponential Smoothing Forecasting Analysis of Product X Sales at Company XYZ Using Double Exponential Smoothing Method,” vol. 22, no. 2, 2023.
- [24] “Lampiran 5”.
- [25] I. Gede, I. Sudipa, and M. Darmawiguna, “BUKU AJAR DATA MINING.” [Online]. Available: <https://www.researchgate.net/publication/377415198>
- [26] C. Herdian, A. Kamila, and I. G. Agung Musa Budidarma, “Studi Kasus Feature Engineering Untuk Data Teks: Perbandingan Label *Encoding* dan *One-Hot Encoding* Pada Metode Linear Regresi,” *Technologia : Jurnal Ilmiah*, vol. 15, no. 1, p. 93, Jan. 2024, doi: 10.31602/tji.v15i1.13457.
- [27] A. P. Sukma Wahyu *et al.*, “PENERAPAN ALGORITMA NAÏVE BAYES CLASSIFICATION UNTUK KLASIFIKASI SENTIMENT TWEET TERHADAP PLATFORM STREAMING ILEGAL,” *Jurnal informasi dan Komputer*, vol. 11, no. 2, p. 2023.
- [28] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, “Teknik K-Fold *Cross Validation* untuk Mengevaluasi Kinerja Mahasiswa,” *Jurnal Algoritma*, vol. 21, no. 1, May 2024, doi: 10.33364/algoritma/v.21-1.1618.