

# Prediksi Target Penjualan Mobil Menggunakan *Decision Tree*

Syahbagus Radithya Haryo Santoso<sup>1</sup>, Punto Adji Bhirawa<sup>2</sup>, Dwi Fibianto<sup>3</sup>, Ghani Fauzan<sup>4</sup>, Basuki Rahmat<sup>5</sup>

<sup>1,2,3,4,5</sup> Program Studi Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur

<sup>1</sup>[22081010255@student.upnjatim.ac.id](mailto:22081010255@student.upnjatim.ac.id)

<sup>2</sup>[22081010231@student.upnjatim.ac.id](mailto:22081010231@student.upnjatim.ac.id)

<sup>3</sup>[22081010321@student.upnjatim.ac.id](mailto:22081010321@student.upnjatim.ac.id)

<sup>4</sup>[22081010258@student.upnjatim.ac.id](mailto:22081010258@student.upnjatim.ac.id)

\*Corresponding author email: [basukirahmat.if@upnjatim.ac.id](mailto:basukirahmat.if@upnjatim.ac.id)

**Abstrak**— Penelitian ini menyoroti pentingnya memahami target pasar dalam penjualan mobil, dengan fokus pada profil demografis konsumen potensial berdasarkan usia dan penghasilan. Menggunakan teknologi *machine learning* dengan metode *decision tree*, penelitian ini bertujuan memprediksi target pasar yang optimal. Data yang digunakan berasal dari platform pembelajaran UdeMy, mencakup berbagai kelompok usia, tingkat penghasilan, dan keinginan individu untuk membeli mobil. Hasil penelitian menunjukkan bahwa *decision tree* dapat membuat prediksi yang akurat, memberikan panduan praktis bagi industri otomotif dalam merumuskan strategi pemasaran dan penjualan. Implementasi analisis ini diharapkan memberikan kontribusi akademis dalam memahami perilaku konsumen terkait pembelian mobil, serta meningkatkan efektivitas kampanye pemasaran dan penjualan. Selain itu, penelitian ini membantu dalam mengembangkan produk dan layanan yang lebih tepat sasaran sesuai dengan kebutuhan konsumen. Kesimpulannya, penelitian ini menunjukkan bahwa metode *decision tree* efektif dalam memprediksi target penjualan mobil berdasarkan profil usia dan penghasilan. Model *decision tree* yang digunakan mencapai akurasi sebesar 91%, yang menunjukkan bahwa metode ini dapat membuat prediksi yang akurat.

**Kata Kunci**— Target Penjualan, Persepsi, Pelanggan, *Deep Learning*, *Decision Tree*.

## I. PENDAHULUAN

Di zaman modern ini, mobil telah menjadi salah satu elemen yang sangat vital dalam aktivitas sehari-hari banyak individu dan keluarga. Mobil tidak hanya berfungsi sebagai alat transportasi yang memudahkan mobilitas, tetapi juga sering dipandang sebagai simbol status sosial dan kenyamanan hidup. Mengingat peran penting yang dimainkan oleh mobil dalam kehidupan masyarakat, memahami target pasar untuk penjualan mobil menjadi sangat krusial bagi produsen dan pemasar. Pengetahuan ini krusial untuk menyusun strategi pemasaran yang efektif dan efisien.

Penelitian ini difokuskan pada identifikasi dan analisis profil demografis konsumen potensial yang memiliki minat untuk

membeli mobil. Tujuan utama dari penelitian ini adalah mengumpulkan dan menganalisis data berdasarkan dua variabel utama yaitu usia dan penghasilan individu. Dengan memahami korelasi antara usia, pendapatan, dan keinginan untuk membeli mobil, penelitian ini bertujuan untuk mengungkap segmen pasar yang paling prospektif bagi industri otomotif. Melalui analisis ini, kita dapat menentukan pada usia berapa individu cenderung memiliki keinginan untuk membeli mobil dan berapa penghasilan yang diperlukan agar seseorang merasa mampu membeli mobil. Selain itu, penelitian ini juga menyoroti bagaimana kombinasi antara usia dan penghasilan mempengaruhi keputusan untuk membeli mobil.

Untuk menentukan target pasar yang optimal, penelitian ini memanfaatkan teknologi *machine learning* dengan menggunakan metode *decision tree*. Metode *decision tree* dipilih karena kemampuannya dalam melakukan prediksi dengan akurasi yang tinggi, yang sangat mendukung perusahaan dalam mengambil keputusan yang akurat (Elhakim, 2024). Prediksi dalam konteks ini adalah teknik untuk memperkirakan suatu nilai di masa depan berdasarkan data historis dan data saat ini. Kemampuan untuk melakukan prediksi yang akurat merupakan aset yang sangat berharga bagi perusahaan atau instansi dalam merumuskan strategi pemasaran dan penjualan yang efektif.

Data yang digunakan dalam penelitian ini diambil dari dataset yang telah tersedia di platform pembelajaran UdeMy. Dataset ini mencakup berbagai kelompok usia dan tingkat penghasilan, yang menyediakan informasi mengenai keinginan individu untuk membeli mobil berdasarkan kondisi keuangan mereka. Analisis data ini diharapkan dapat memberikan wawasan yang jelas tentang segmen pasar yang paling prospektif untuk penjualan mobil dan membantu para pemasar dalam merancang strategi pemasaran yang lebih tepat sasaran.

Dalam menghadapi tantangan dalam hal teknologi, diperlukan alat untuk menganalisis data penjualan mereka (Jeffry et al, 2023). Hasil dari analisis ini diharapkan tidak hanya memberikan kontribusi akademis dalam memahami perilaku konsumen terkait pembelian mobil, tetapi juga menawarkan panduan praktis bagi para pelaku industri otomotif. Dengan

mengidentifikasi dan menjangkau segmen pasar yang paling relevan, produsen mobil dapat meningkatkan efektivitas kampanye pemasaran mereka dan pada akhirnya dapat meningkatkan penjualan. Selain itu, hasil analisis ini dapat dimanfaatkan untuk mengembangkan produk dan layanan yang lebih cocok dengan kebutuhan dan preferensi konsumen, sehingga menciptakan nilai tambah bagi perusahaan dan konsumen.

Dengan memahami target pasar secara mendalam, distributor dan produsen mobil dapat menyesuaikan pendekatan mereka untuk memenuhi kebutuhan dan harapan konsumen. Penelitian ini diharapkan mampu memberikan wawasan yang berharga tentang faktor-faktor yang mempengaruhi keputusan pembelian mobil sehingga dapat membantu dalam penyusunan strategi bisnis yang lebih efektif (Setiawan, 2024). Dengan demikian, diharapkan penelitian ini akan berkontribusi pada peningkatan penjualan mobil serta pengembangan produk dan layanan yang lebih unggul, sehingga pada akhirnya akan meningkatkan kepuasan konsumen dan keberhasilan bisnis di industri otomotif.

## II. METODE

Pada penelitian kali ini kami menggunakan metode *Decision Tree* untuk memprediksi target penjualan mobil dengan beberapa proses yang kami lakukan untuk mencapai hasil akhir yang diinginkan, metodenya antara lain:

- Proses Pertama **Exploratory Data**  
*Exploratory Data Analysis* memungkinkan *analyst* memahami isi data yang digunakan mulai dari distribusi, frekuensi, korelasi, dan lainnya. Pada praktik di proses pertama, *curiosity* sangatlah penting, pemahaman konteks data juga diperhatikan karena akan menjawab masalah-masalah dasar (Data Folks Indonesia, 2019).
- Proses Kedua **Pre-Processing Data**  
Proses *pre-processing data* merupakan sebuah metode awal yang dilakukan sebelum tahap pemrosesan data atau bisa dibilang pada metode ini kita mengolah sebuah data yang masih mentah agar menjadi data yang lebih baik (Suripto et al, 2022).
- Proses Ketiga **Split Data**.  
Proses ketiga yaitu *split data* merupakan sebuah proses yang mana data yang kita miliki akan dibagi menjadi dua bagian, yaitu data *training* dan data *testing*, dengan komposisi data *training* 75% sedangkan data *testing* 25%. Fungsi data *training* yaitu untuk melatih sebuah model agar bisa mengidentifikasi pola yang relevan, sedangkan data *testing* merupakan implementasi dari pola-pola yang sudah dilatih pada data *training* (Sugumonrong and Gultom, 2023).
- Proses Keempat **Feature Scaling**  
Penskalaan data memainkan peran penting dalam pra-pemrosesan data yang mempengaruhi kinerja algoritma pembelajaran mesin. Penelitian ini bertujuan untuk menganalisis dampak teknik normalisasi min-max dan

standarisasi (normalisasi nol-mean) terhadap kinerja algoritma pembelajaran mesin dalam skala dan jangkauan yang sama sehingga bisa membantu untuk mendapatkan hasil akhir (Ambarwari et al, 2022).

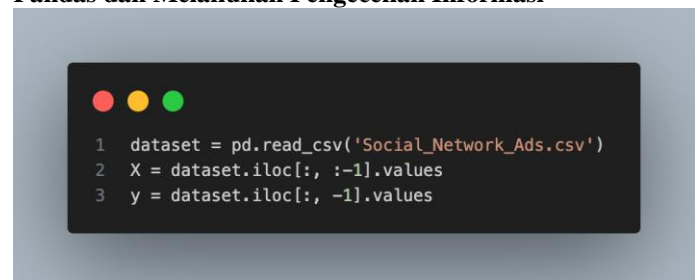
- Proses Kelima **Confusion Matrix**  
*Confusion matrix* adalah tabel yang digunakan untuk menentukan kinerja model. *Confusion matrix* memvisualisasikan dan merangkum kinerja algoritma klasifikasi (Awujoola et al, 2022).

- Proses Keenam **Visualisasi Data**  
Visualisasi data adalah proses menggunakan elemen visual seperti diagram, grafik, atau peta untuk merepresentasikan data. Visualisasi data menerjemahkan yang kompleks, bervolume tinggi, atau numerik menjadi representasi visual yang lebih mudah diproses. Alat visualisasi data dapat meningkatkan dan otomatisasi proses komunikasi visual untuk mendapatkan akurasi dan detail. Anda dapat menggunakan representasi visual untuk mengekstraksi wawasan yang dapat ditindaklanjuti dari data mentah (Muktiwijaya et al. 2024)

## III. HASIL DAN PEMBAHASAN

Sumber data dalam penelitian ini diambil dari data penjualan mobil mewah suv yang terdiri dari dua kolom sebagai variabel *independen* dan satu kolom sebagai variabel *dependen*. *Dataset* dapat diakses melalui [link](#) berikut [Social Network Ads.csv](#)

### 1. Melakukan Import Data Menggunakan Library Pandas dan Melakukan Pengecekan Informasi

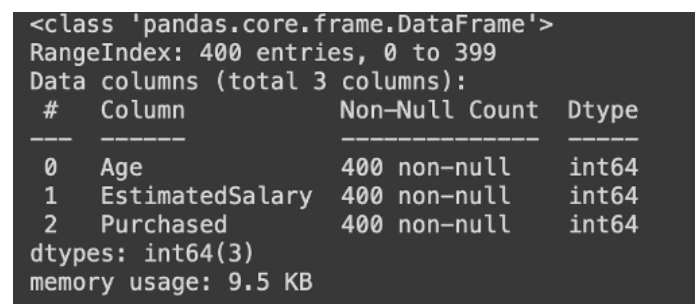


```

1 dataset = pd.read_csv('Social_Network_Ads.csv')
2 X = dataset.iloc[:, :-1].values
3 y = dataset.iloc[:, -1].values

```

Gbr. 1 Gambar code mendefinisikan dataset



```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 3 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Age              400 non-null   int64
1   EstimatedSalary  400 non-null   int64
2   Purchased        400 non-null   int64
dtypes: int64(3)
memory usage: 9.5 KB

```

Gbr. 2 Gambar kolom beserta tipe data

*Output* akan menampilkan informasi tentang *dataset*, termasuk jumlah baris dan kolom, serta tipe data dari setiap kolom. Jika semua kolom berisi tipe data *integer*, maka *output* akan mencantumkan informasi tersebut.

## 2. Memisahkan Dataset Menjadi *Training Set* dan *Test Set*

```

1 from sklearn.model_selection import train_test_split
2 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 0)

```

Gbr. 3 Contoh gambar membagi data *train* dan data *test*

Pemisahan *dataset* menjadi *training set* dan *test set* karena keduanya merupakan kebutuhan utama dalam *machine learning* agar dapat diproses sesuai dengan rancangannya (MR, 2023). *Training set* digunakan untuk membuat model *machine learning* sedangkan *test set* digunakan untuk menguji kebenaran dari model yang dibuat (Kukuh, 2018).

## 3. Feature Scaling

Teknik yang digunakan dalam pra-pemrosesan data untuk menormalkan rentang fitur dalam data. Ini sangat penting dalam banyak algoritma *machine learning* karena beberapa algoritma sensitif terhadap skala data. Berikut adalah beberapa kegunaan utama dari *feature scaling* (Shivanipickl, 2023):

- Meningkatkan konvergensi algoritma
- Menghindari fominasi fitur tertentu
- Meningkatkan kinerja model
- Mempercepat proses pelatihan

```

1 from sklearn.preprocessing import StandardScaler
2 sc = StandardScaler()
3 X_train = sc.fit_transform(X_train)
4 X_test = sc.transform(X_test)

```

Gbr. 4 Contoh gambar code *feature scaling*

Berikut adalah data setelah melewati proses *feature scaling*.

```

[[ 0.58164944 -0.88670699]
 [-0.60673761 1.46173768]
 [-0.01254409 -0.5677824 ]
 [-0.60673761 1.89663484]
 [ 1.37390747 -1.40858358]
 [ 1.47293972 0.99784738]
 [ 0.08648817 -0.79972756]
 [-0.01254409 -0.24885782]
 [-0.21060859 -0.5677824 ]
 [-0.21060859 -0.19087153]
 [-0.30964085 -1.29261101]
 [-0.30964085 -0.5677824 ]
 [ 0.38358493 0.09905991]
 [ 0.8787462 -0.59677555]
 [ 2.06713324 -1.17663843]
 [ 1.07681071 -0.13288524]
 [ 0.68068169 1.78066227]
 [-0.70576986 0.56295021]
 [ 0.77971394 0.35999821]
 [ 0.8787462 -0.53878926]
 [-1.20093113 -1.58254245]
 [ 2.1661655 0.93986109]
 [-0.01254409 1.22979253]

```

Gbr. 5 Contoh output pada data *X\_train* setelah *feature scaling*

## 4. Melatih Model *Decision Tree Classification Model* Pada *Training Set*

Model akan memulai dengan seluruh *dataset* dan mencoba membagi data menjadi dua kelompok berdasarkan kriteria tertentu. Dalam proses ini menggunakan entropi untuk mengukur *impurity*. Proses pembagian ini terus berlanjut secara rekursif, dengan membagi *dataset* menjadi *subset* yang lebih kecil pada setiap level hingga berhenti pada kondisi tertentu, seperti jumlah maksimum kedalaman pohon atau jumlah minimum sampel dalam *node*. Entropi digunakan untuk mengukur ketidakmurnian dari *node*. *Node* yang lebih murni memiliki entropi yang lebih rendah. Model akan memilih *split* yang mengurangi entropi secara signifikan.

## 5. Memprediksi Hasil *Test Set*

*Predicting the test set results* adalah langkah penting dalam proses pengembangan model *machine learning*. Ini melibatkan model yang telah dilatih pada *train set* untuk membuat prediksi pada data yang baru atau belum pernah dilihat sebelumnya.

Tujuan dari *Predicting the Test Set Results*:

1. Evaluasi kinerja model
2. Mengukur generalisasi
3. Identifikasi kesalahan dan perbaikan
4. Validasi model

## 6. Confusion Matrix

*Confusion matrix* berfungsi untuk mengevaluasi kinerja model klasifikasi. Tabel *confusion matrix* memungkinkan kita untuk memvisualisasikan kinerja algoritma klasifikasi, terutama ketika kita memiliki lebih dari dua kelas. Setiap kolom dari matriks ini mewakili prediksi dari model, sedangkan setiap baris mewakili *instance* aktual dari kelas (Narkhede, 2024).

		Actual class	
		P	N
Predicted class	P	TP	FP
	N	FN	TN

Gbr. 6 Contoh gambar tabel *confusion matrix*  
Sumber: Tiwari (2022)

Komponen *Confusion Matrix* (Diez, 2018):

- **True Negative (TN)**: Jumlah *instance* negatif yang diprediksi dengan benar sebagai negatif.
- **False Positive (FP)**: Jumlah *instance* negatif yang diprediksi secara salah sebagai positif (*false alarm*).
- **False Negative (FN)**: Jumlah *instance* positif yang diprediksi secara salah sebagai negatif (*miss*).

- **True Positive (TP):** Jumlah *instance* positif yang diprediksi dengan benar sebagai positif.

Pada artikel penelitian ini menggunakan Akurasi (*Accuracy*). Pada kode program menunjukkan Akurasi yang cukup tinggi sebanyak 91%.

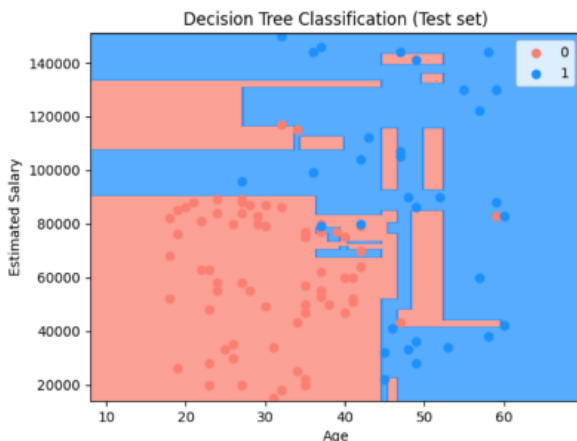
```
[[62  6]
 [ 3 29]]
0.91
```

Gbr. 7 Output dari *confusion matrix*

## 7. Visualisasi Data



Gbr. 8 Visualisasi *train set*



Gbr. 8 Gambar visualisasi *test set*

Pada langkah ini, menampilkan visualisasi data yang diperoleh dari *training set* dan *test set*. Visualisasi membantu kita memahami cara model pembelajaran mesin dan memisahkan kelas-kelas yang berbeda dalam ruang fitur. Hal ini memungkinkan kita untuk melihat secara langsung bagaimana model mempelajari pola dari data pelatihan.

## IV. KESIMPULAN

Penelitian ini menunjukkan bahwa metode *decision tree* efektif dalam memprediksi target penjualan mobil berdasarkan profil usia dan penghasilan. Dengan menganalisis data dari platform pembelajaran Udemy, model ini berhasil mengidentifikasi segmen pasar yang prospektif, memberikan panduan praktis bagi industri otomotif dalam menyusun strategi pemasaran dan penjualan mobil suv. Implementasi metode ini membantu dalam memahami perilaku konsumen, sehingga produsen dan pemasar dapat menyesuaikan pendekatan mereka untuk memenuhi kebutuhan konsumen. Model *decision tree* yang digunakan mencapai akurasi sebesar 91%, yang menunjukkan bahwa metode ini dapat membuat prediksi yang akurat. Dengan demikian, hasil penelitian ini berkontribusi pada peningkatan efektivitas kampanye pemasaran, pengembangan produk yang lebih tepat sasaran, dan peningkatan kepuasan serta keberhasilan bisnis di industri otomotif.

## REFERENSI

- [1] Ambarwari, Agus, Qadhli J. Adrian, and Yeni Herdiyeni. 2022. Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman. *RESTI (Rekayasa Sistem dan Teknologi Informasi)* 4 (1): 7. 10.29207/resti.v4i1.1517.
- [2] Awujoola, Olalekan J., dkk. 2023. *Data Science for Genomics*. N.p.: ACADEMIC PRESS. <https://doi.org/10.1016/B978-0-323-98352-5.00013-6>
- [3] Data Folks Indonesia. 2019. Memahami Data Dengan Exploratory Data Analysis | by AC | Data Folks Indonesia. Medium. <https://medium.com/data-folks-indonesia/memahami-data-dengan-exploratory-data-analysis-a53b230cce84>.
- [4] Diez, Pablo. 2018. *Smart Wheelchairs and Brain-Computer interfaces*. N.p.: ACADEMIC PRESS. <https://doi.org/10.1016/B978-0-12-812892-3.00001-7>.
- [5] Elhakim, Puja Naufal Humam. 2024. Penerapan algoritma decision tree untuk menentukan kelayakan pemberian kredit usaha. Tesis, Teknik Informatika. Bandung:UIN Sunan Gunung Djati. <https://digilib.uinsgd.ac.id/86725/>
- [6] Jeffry, Nurdyansa, and Syahrul Usman. 2023. "TRANSFORMASI DIGITAL DALAM INDUSTRI KERAJINAN RUMAHAN MELALUI PENGGUNAAN TEKNOLOGI MACHINE LEARNING." *JMM (Jurnal Masyarakat Mandiri)* 7, no. 6 (Desember): 10. <https://doi.org/10.31764/jmm.v7i6.19209>.
- [7] Kukuh, R. 2018. "Memisah Dataset Menjadi Training-Set dan Test-Set." Medium. <https://medium.com/@rkukuh/memisah-dataset-menjadi-training-set-dan-test-set-19b45dd52f6d>.
- [8] MR, Salsabila. 2023. "3 Tipe Proses Dataset dalam Mechine Learning." DQLab. <https://dqlab.id/3-tipe-proses-dataset-dalam-machine-learning#:~:text=Data%20Training%20dan%20Data%20Testing,Machine%20Learning%20sesuai%20dengan%20rancangannya>.
- [9] Muktiwijaya, Aldi W., dkk. 2024. Visualisasi Data tentang Persepsi Mahasiswa terhadap Penggunaan Qris di Provinsi Daerah Istimewa Yogyakarta. *Innovative: Journal Of Social Science Research* 4 (4): 16. <https://doi.org/10.31004/innovative.v4i4.12597>.
- [10] Narkhede, Sarang. n.d. Understanding Confusion Matrix | by Sarang Narkhede. Towards Data Science. Accessed July 17, 2024. <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>.
- [11] Setiawan, Zunan., dkk. 2024. *Buku Ajar Perilaku Konsumen*. N.p.: PT. Sonpedia Publishing Indonesia Shivanipickl. 2023. What is Feature Scaling and Why Does Machine Learning Need It? Medium. <https://medium.com/@shivanipickl/what-is-feature-scaling-and-why-does-machine-learning-need-it-104e4edbb1c9>.
- [12] Shivanipickl. 2023. What is Feature Scaling and Why Does Machine Learning Need It? Medium. <https://medium.com/@shivanipickl/what-is-feature-scaling-and-why-does-machine-learning-need-it-104e4edbb1c9>.

- [13] Sugumonrong, Darwin P., and Diana A. Gultom. 2023. Perbandingan Metode Moving Average (MA) Dan Neural Network yang Berbasis Algoritma Backpropagation Dalam Prediksi Harga Saham. *INFORMATION SYSTEM DEVELOPMENT* 3, no. 2 (Juli): 9. [https://d1wqtxts1xzle7.cloudfront.net/81443083/193-libre.pdf?1646019934=&response-content-disposition=inline%3B+filename%3DPerbandingan\\_Metode\\_Moving\\_Average\\_Ma\\_Da.pdf&Expires=1723024848&Signature=JRsi-EqH0Z2sgW0bk8bIVyTTVBXZwcWY~fXkhrIElWJ7K399540pRBBFYC4](https://d1wqtxts1xzle7.cloudfront.net/81443083/193-libre.pdf?1646019934=&response-content-disposition=inline%3B+filename%3DPerbandingan_Metode_Moving_Average_Ma_Da.pdf&Expires=1723024848&Signature=JRsi-EqH0Z2sgW0bk8bIVyTTVBXZwcWY~fXkhrIElWJ7K399540pRBBFYC4).
- [14] Suropto, Rr N. Rahmanita, and Ajeng S. Kirana. 2022. Teknik pre-processing dan classification dalam data science. BINUS UNIVERSITY. <https://mie.binus.ac.id/2022/08/26/teknik-pre-processing-dan-classification-dalam-data-science/>.
- [15] Tiwari, Ashish. (2022). *Artificial Intelligence and Machine Learning for EDGE Computing*. N.p.: ACADEMIC PRESS. <https://doi.org/10.1016/B978-0-12-824054-0.00026-5>