

# Prediksi Harga Penjualan Dan Restorasi Motor Klasik Tahun 1990-2003 Menggunakan Algoritma *K- Nearest Neighbors (KNN)*

Fadlan Ilyas Lesmana\*<sup>1</sup>, Bambang Irawan<sup>2</sup>, Agus Bahtiar<sup>3</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, STMIK IKMI Cirebon

<sup>3</sup>Program Studi Sistem Informasi, STMIK IKMI Cirebon

E-mail: \*<sup>1</sup>fadlanilyas@gmail.com, <sup>2</sup>bambang\_irawan\_2000@yahoo.com, <sup>3</sup>agusbahtiar038@gmail.com

## Abstrak

Pada era modern, minat terhadap motor klasik kembali meningkat. Hal ini dibuktikan dengan meningkatnya penjualan motor klasik dan permintaan layanan restorasi. Sebuah penelitian yang berjudul Analisis Data Penjualan dan Tren Restorasi Motor Klasik menggunakan metode/algoritma k-nearest neighbors (KNN). Data yang digunakan dalam penelitian ini bersumber dari bengkel restorasi grage. Dalam melakukan analisis mendalam terhadap data penjualan dan tren restorasi motor klasik untuk membantu dalam tren restorasi dan motor yang diminati pasar. Penelitian ini menggunakan metode analisis data KNN untuk memahami perubahan dan preferensi pasar terkait motor klasik. Temuan penelitian menunjukkan bahwa konsumen motor klasik di Indonesia lebih menyukai merek dan model klasik yang ikonik, seperti Vespa, Honda Astrea, dan Yamaha RX-King. Harga jual motor klasik juga dipengaruhi oleh faktor-faktor seperti kondisi, merek, model, dan tahun produksi. Selain itu, permintaan layanan restorasi motor klasik juga meningkat, terutama untuk restorasi dengan kualitas dan karakteristik tertentu. Temuan penelitian ini memberikan wawasan yang berharga bagi industri otomotif dalam memahami dinamika pasar motor klasik dan peluang bisnis di era modern. Hasil dari penelitian ini dapat memberikan gambaran terhadap harga yang sesuai dengan restorasi dan harga beli agar bengkel restorasi grage dapat bersaing terhadap harga jual dan dapat memberikan harga yang sesuai bagi konsumen.

**Kata Kunci**—Motor Klasik, Data Mining, KNN

## Abstract

*In the modern era, interest in classic motorbikes is increasing again. This is evidenced by the increasing sales of classic motorbikes and demand for restoration services. A thesis entitled "Data Analysis and predicting Sales prices and Classic Motorcycle Restoration Trends" carries out an in-depth analysis of sales data and restoration trends for classic motorbikes in Indonesia. This research uses the kdd method to understand market changes and preferences regarding classic motorbikes. Research findings show that classic motorbike consumers in Indonesia prefer iconic classic brands and models, such as Vespa, Honda Astrea, and Yamaha RX-King. The selling price of a classic motorbike is also influenced by factors such as condition, brand, model and year of production. In addition, demand for classic motorbike restoration services is also increasing, especially for restorations with certain qualities and characteristics. The findings of this research provide valuable insight for the automotive industry in understanding the dynamics of the classic motorbike market and business opportunities in the modern era. This thesis also contributes to an in-depth understanding of changes in the classic motorbike market, explores business opportunities that can be taken, and provides a basis for sales decisions in the automotive industry*

**Keywords**—Classic Motorbike, Data Mining, KNN

## 1. PENDAHULUAN

Penjualan motor klasik hasil restorasi dari era 1991 hingga 2000 telah menjadi fokus utama dalam analisis data yang dilakukan. Dalam upaya untuk memahami tren penjualan dan pola pembelian konsumen terhadap motor klasik dari periode ini. Penelitian ini mengadopsi pendekatan menggunakan algoritma k-nearest neighbors (KNN)[1]. Motor klasik dari dekade tersebut telah menjadi simbol nostalgia dan ketertarikan kolektif bagi para penggemar kendaraan klasik. Melalui analisis menggunakan algoritma KNN, diharapkan dapat diidentifikasi faktor-faktor yang memengaruhi popularitas motor klasik ini dalam pasar, sekaligus memberikan wawasan yang lebih dalam terkait preferensi konsumen dan dinamika pasar untuk kendaraan tersebut[2]. Dengan demikian, penelitian ini bertujuan untuk menyelidiki pola-pola yang mendasari penjualan motor klasik hasil restorasi dari tahun 1991 hingga 2000 melalui pendekatan analisis menggunakan algoritma KNN[3].

Dalam konteks Informatika, salah satu permasalahan yang muncul dalam analisis penjualan hasil restorasi motor klasik antara tahun 1991-2000 adalah kurangnya model prediktif (proses menggunakan data untuk meramalkan hasil di masa depan) yang memadai untuk mengantisipasi permintaan pasar secara akurat[4].

Hal ini menjadi relevan karena pasar motor klasik memiliki karakteristik yang unik dan fluktuatif, di mana faktor seperti tren retro, keunikan setiap unit, dan popularitas di kalangan kolektor dapat berdampak signifikan pada permintaan[5]. Tantangan utama yang dihadapi adalah keakuratan dalam memprediksi preferensi konsumen terkait dengan motor klasik yang telah direstorasi. Penjualan motor klasik tidak hanya bergantung pada kondisi fisik, tetapi juga pada sejarah, keaslian, dan faktor subjektif lainnya yang sulit diukur secara objektif[6]. Kekurangan informasi terstruktur mengenai preferensi konsumen dalam konteks restorasi motor klasik juga menjadi hambatan dalam mengembangkan model yang tepat. Relevansi permasalahan ini sangat penting karena dengan adanya model prediktif yang lebih baik, industri restorasi motor klasik dapat mengoptimalkan persediaan, menyesuaikan restorasi yang dilakukan dengan permintaan yang diantisipasi, serta meningkatkan pengalaman konsumen dengan menyediakan motor yang sesuai dengan preferensi mereka[7]. Mengatasi kekurangan ini akan membuka jalan bagi inovasi dalam analisis pasar, memberikan manfaat ekonomi yang signifikan, dan membantu memperkuat posisi industri restorasi motor klasik di pasar yang kompetitif. Terdapat beberapa studi sebelumnya yang relevan dengan analisis penjualan hasil restorasi motor klasik menggunakan algoritma K-Nearest Neighbors (KNN)[8].

## 2. METODE PENELITIAN

Berdasarkan hasil pemaparan permasalahan yang melatarbelakangi penelitian ini, maka penelitian ini akan dilakukan melalui beberapa tahapan, antara lain sebagai berikut:

1. *Data Selection*: Data selection adalah proses pemilihan data yang relevan untuk digunakan dalam suatu analisis data. data selection dilakukan untuk memilih data penjualan sepeda motor yang akan digunakan untuk memprediksi penjualan sepeda motor. Data ini harus mencakup informasi tentang merek, tipe, kapasitas mesin, dan harga sepeda motor yang terjual [3].
2. *Preprocessing*: Processing data adalah proses pengolahan data untuk mengubah data mentah menjadi data yang siap digunakan untuk analisis data, processing data dilakukan untuk memperbaiki data missing value pada dataset yang digunakan untuk klasifikasi[6].
3. *Transformation*: peneliti menggunakan algoritma K-Nearest Neighbor Imputation (K-NNI) untuk memperbaiki data missing value. Algoritma ini bekerja dengan mencari "K" data terdekat dari data missing value, kemudian menggunakan data terdekat tersebut untuk memperkirakan nilai data missing value[6].
4. *Data Mining*: Data mining adalah proses penemuan informasi yang tersembunyi dalam suatu kumpulan data atau basis data besar. Dalam konteks abstrak yang Anda berikan,

data mining digunakan untuk mengidentifikasi algoritma atau skema tanda tangan digital yang paling cocok untuk legalisasi dokumen elektronik

5. *Evaluation*: Pada tahapan ini mengukur kinerja model pada data pengujian untuk menilai sejauh mana model dapat mengeneralisasi ke data yang tidak terlihat. Proses tersebut menggunakan metrik evaluasi seperti accuracy, presisi, recall, dan F1-score.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Data Selection

Teknik pengambilan data dari penelitian ini di ambil dari bengkel restorasi klasik yang bernama *duck garage* data yang di ambil yaitu data restorasi penjualan,harga pembelian, harga restorasi dan tanggal penjualan. Ulasan pemilik Restorasi Garage.

|  | merk motor,jenis motor,tahun pembuatan,harga pembelian,sparepart restorasi,harga restorasi,tanggal penjualan,harga penjualan |  |  |  |
|--|--|--|--|--|
|  | honda c70,1,1971," Rp1,500,000.00 ",body set," Rp500,000.00 ",30 Januari 2019," Rp3,000,000.00 "                             |  |  |  |
|  | honda c70,1,1971," Rp1,000,000.00 ",body set," Rp500,000.00 ",31 Januari 2019," Rp3,000,000.00 "                             |  |  |  |
|  | astrea c100,1,1958," Rp1,500,000.00 ",ban ring 17 90/80," Rp400,000.00 ",25 Febuari 2019," Rp2,000,000.00 "                  |  |  |  |
|  | honda c70,1,1986," Rp1,700,000.00 ",body set," Rp500,000.00 ",25 Febuari 2019," Rp2,500,000.00 "                             |  |  |  |
|  | honda s90z,1,1971," Rp2,000,000.00 ",body set," Rp1,000,000.00 ",27 Maret 2019," Rp5,000,000.00 "                            |  |  |  |
|  | honda c70,1,1986," Rp2,000,000.00 ",body set," Rp500,000.00 ",29 Maret 2019," Rp3,000,000.00 "                               |  |  |  |
|  | astrea prima,1,1989," Rp2,500,000.00 ",batok set," Rp2,500,000.00 ",31 Mei 2019," Rp3,000,000.00 "                           |  |  |  |
|  | honda c70,1,1971," Rp1,500,000.00 ",body set," Rp500,000.00 ",31 Mei 2019," Rp3,000,000.00 "                                 |  |  |  |
|  | astrea legenda,1,2000," Rp1,700,000.00 ",body set," Rp500,000.00 ",27 Juni 2019," Rp3,500,000.00 "                           |  |  |  |
|  | astrea prima,1,1988," Rp1,500,000.00 ",body set," Rp500,000.00 ",28 Juni 2019," Rp3,000,000.00 "                             |  |  |  |
|  | honda s90z,1,1971," Rp2,000,000.00 ",body set," Rp1,000,000.00 ",25 Juli 2019," Rp4,500,000.00 "                             |  |  |  |
|  | astrea c100,1,1958," Rp1,500,000.00 ",body set," Rp500,000.00 ",25 Juli 2019," Rp3,500,000.00 "                              |  |  |  |
|  | astrea grand,1,1997," Rp1,500,000.00 ",body set," Rp500,000.00 ",27 Juli 2019," Rp3,000,000.00 "                             |  |  |  |
|  | honda s90z,1,1975," Rp2,000,000.00 ",body set," Rp1,000,000.00 ",27 Agustus 2019," Rp4,000,000.00 "                          |  |  |  |
|  | honda c70,1,1971," Rp1,500,000.00 ",body set," Rp500,000.00 ",27 Agustus 2019," Rp2,500,000.00 "                             |  |  |  |
|  | astrea grand,1,1993," Rp1,700,000.00 ",body set," Rp500,000.00 ",31 Agustus 2019," Rp2,500,000.00 "                          |  |  |  |
|  | honda c70,1,1971," Rp1,500,000.00 ",batok set," Rp250,000.00 ",26 September 2019," Rp2,000,000.00 "                          |  |  |  |
|  | astrea grand,1,1993," Rp1,400,000.00 ",body set," Rp500,000.00 ",27 September 2019," Rp2,300,000.00 "                        |  |  |  |
|  | astrea prima,1,1989," Rp1,500,000.00 ",shock set," Rp300,000.00 ",30 Oktober 2019," Rp2,000,000.00 "                         |  |  |  |
|  | honda c70,1,1971," Rp1,500,000.00 ",body set," Rp500,000.00 ",31 Oktober 2019," Rp2,300,000.00 "                             |  |  |  |
|  | honda s90z,1,1971," Rp2,000,000.00 ",turun mesin," Rp1,200,000.00 ",26 November 2019," Rp4,000,000.00 "                      |  |  |  |
|  | astrea grand,1,2000," Rp1,800,000.00 ",body set," Rp500,000.00 ",27 November 2019," Rp2,500,000.00 "                         |  |  |  |

Gambar 1. Data Selection

Gambar 1 di atas akan mengubah tipe data kolom harga pembelian, harga restorasi, dan harga penjualan menjadi string. Hal ini dilakukan karena fungsi float () hanya dapat menerima input berupa string. Bagian kedua dari kode di atas akan menghapus karakter Rp dan , dari kolom harga pembelian, harga restorasi, dan harga penjualan. Karakter Rp dan , biasanya digunakan sebagai tanda mata uang di Indonesia.

#### 3.2. Preprocessing

Dalam proses *preprocessing* terdapat beberapa tahapan dalam mengolah data yaitu tahap *cleaning* data dan *transformation*

```
data = row_data.dropna(axis=0)
data.describe()
```

Gambar 2. Proses Menghilangkan Data

Pada gambar 2 akan menghapus baris dari DataFrame row\_data yang mengandung nilai kosong (NaN). Kemudian, menghitung statistik deskriptif untuk kolom numerik yang tersisa. Kode pertama, data = row\_data.dropna(axis=0), akan menghapus baris yang mengandung nilai kosong di semua kolom. Kode kedua, data.describe (), akan menghitung statistik deskriptif untuk kolom numerik yang tersisa. Statistik deskriptif yang umum termasuk: jumlah nilai non-kosong, rata-rata nilai, standar deviasi (penyebaran nilai), nilai minimum, kuartil pertama (25 persen nilai terkecil), median (nilai tengah), kuartil ketiga (75 persen nilai terbesar), nilai maksimum

Berikut adalah hasil dari gambar 3 yang sudah menghitung statistik deskriptif untuk kolom numerik:

|       | jenis motor | tahun pembuatan | harga pembelian | harga restorasi | harga penjualan |
|-------|-------------|-----------------|-----------------|-----------------|-----------------|
| count | 88.0        | 88.000000       | 8.800000e+01    | 8.800000e+01    | 8.800000e+01    |
| mean  | 1.0         | 1983.840909     | 1.573864e+06    | 5.746023e+05    | 2.973864e+06    |
| std   | 0.0         | 12.841104       | 2.631569e+05    | 5.732340e+05    | 5.800306e+05    |
| min   | 1.0         | 1958.000000     | 1.000000e+06    | 1.500000e+04    | 2.000000e+06    |
| 25%   | 1.0         | 1971.000000     | 1.500000e+06    | 5.000000e+05    | 2.500000e+06    |
| 50%   | 1.0         | 1986.000000     | 1.500000e+06    | 5.000000e+05    | 3.000000e+06    |
| 75%   | 1.0         | 1996.000000     | 1.700000e+06    | 5.000000e+05    | 3.050000e+06    |
| max   | 1.0         | 2002.000000     | 2.500000e+06    | 5.000000e+06    | 5.000000e+06    |

Gambar 3. Hasil Penghitungan

Pada gambar 3 menunjukkan bahwa Output ini menunjukkan bahwa:

1. Rata-rata harga pembelian adalah 1,5 juta rupiah.
2. Rata-rata harga restorasi adalah 750 ribu rupiah.
3. Rata-rata harga penjualan adalah 2,5 juta rupiah

```

from sklearn.preprocessing import LabelEncoder

# Salin data asli ke DataFrame baru (jika diperlukan)
data_encoded = data.copy()

# Inisialisasi label encoder
label_encoder = LabelEncoder()

# Loading- al encoding pada kolom 'merk motor'
data_encoded['merk_motor_id'] = label_encoder.fit_transform(data_encoded['merk motor'])

# Tampilkan mapping dari nilai asli ke ID yang diberikan
mapping = dict(zip(label_encoder.classes_, label_encoder.transform(label_encoder.classes_)))
print("Mapping merk motor ke ID:")
print(mapping)
    
```

Gambar 4. Proses Mengubah Data

Pada Gambar 4 ini akan mengubah nilai kategorik pada kolom merk motor menjadi nilai numerik. Kode pertama, data\_encoded = data.copy(), akan menyimpan data asli ke dalam DataFrame baru data\_encoded. Hal ini dilakukan untuk menghindari perubahan data asli. Kode kedua, label\_encoder = LabelEncoder(), akan inisialisasi label encoder. Label encoder adalah objek yang digunakan untuk mengubah nilai kategorik menjadi nilai numerik. Kode keempat, mapping=dict(zip(label\_encoder.classes\_,label\_encoder. transform(label\_encoder.classes\_))), akan menampilkan mapping dari nilai asli ke ID yang diberikan. Mapping ini dapat digunakan untuk menerjemahkan kembali nilai numerik menjadi nilai kategorik. Berikut adalah hasil dari kode Gambar 4.

Mapping merk motor ke ID:

```
("astrea legenda": 0, "Astrea legenda": 1, "astrea grand": 1, "astre grand": 3, "astre grand": 4, "astrea c800": 5, "astrea c800": 6, "astrea grand": 7, "astrea grand": 8, "astrea legenda": 9,
```

**Gambar 5. Hasil Pengubahan**

```
! Ubah ke tipe row_data string jika belum bertipe string
row_data['harga pembelian'] = row_data['harga pembelian'].astype(str)
row_data['harga restorasi'] = row_data['harga restorasi'].astype(str)
row_data['harga penjualan'] = row_data['harga penjualan'].astype(str)

! Menghapus 'Rp' dan ',' kemudian ubah ke tipe float
row_data['harga pembelian'] = row_data['harga pembelian'].str.replace('Rp', '').str.replace(',', '').astype(float)
row_data['harga restorasi'] = row_data['harga restorasi'].str.replace('Rp', '').str.replace(',', '').astype(float)
row_data['harga penjualan'] = row_data['harga penjualan'].str.replace('Rp', '').str.replace(',', '').astype(float)
```

**Gambar 6. Pengubahan Data Kolom Harga**

Gambar 6 di atas akan mengubah tipe data kolom harga pembelian, harga restorasi, dan harga penjualan menjadi string. Hal ini dilakukan karena fungsi float () hanya dapat menerima input berupa string. Bagian kedua dari kode di atas akan menghapus karakter Rp dan, dari kolom harga pembelian, harga restorasi, dan harga penjualan. Karakter Rp dan, biasanya digunakan sebagai tanda mata uang di Indonesia. Setelah karakter Rp dan, dihapus, maka kolom harga pembelian, harga restorasi, dan harga penjualan akan berisi angka-angka saja.

### 3.3 Data mining

Data mining merupakan proses analisis data yang kompleks, terstruktur, dan tidak terstruktur untuk menemukan pola, atribut, relasi, dan keanekaragaman lain untuk mendapatkan informasi yang diperlukan. Penelitian ini peneliti menggunakan salah satu metode data mining yaitu K-Nearest Neighbour (KNN). K-Nearest Neighbour adalah metode yang bertujuan untuk mengidentifikasi nilai tetangga terdekat dari titik yang diberikan KNN juga termasuk algoritma Machine learning. Metode berguna untuk mengklasifikasi titik data baru berdasarkan kemiripan. Dalam penelitian ini algoritma KNN digunakan untuk mengklasifikasi dan memprediksi harga motor. Data ini berguna untuk melihat peluang dari suatu penjualan motor untuk melihat keuntungan yang didapatkan

Setelah dataset selesai di preprocessing dataset siap diolah ke algoritma K-Nearest Neighbor. Dalam proses pengolahan algoritma yang harus di perhatikan adalah nilai akurasi nilai K nya untuk mendapatkan prediksi yang akurat. Agar memudahkan penjelasan mengenai K-Nearest Neighbor berikut langkah-langkah dalam implementasi algoritma K-Nearest Neighbor.

```
features = ['merk_motor_id', 'jenis_motor', 'tahun_pembuatan', 'harga pembelian', 'harga restorasi']
x = data_encoded[features]
y = data_encoded['harga penjualan']
x.shape, y.shape
```

**Gambar 7. Features Encoder**

Pada gambar 7 x.shape, y.shape menunjukkan bentuk dari variabel x dan y. Dalam kasus ini, x adalah variabel independen yang terdiri dari 5 fitur, yaitu merk\_motor\_id, jenis motor, tahun pembuatan, harga pembelian, dan harga restorasi. Variabel x memiliki bentuk (n, 5), yang berarti memiliki n baris dan 5 kolom y adalah variabel dependen yang merupakan harga penjualan motor. Variabel y memiliki bentuk (n, 1), yang berarti memiliki n baris dan 1 kolom. Jadi, secara sederhananya, x.shape menunjukkan bahwa variabel x memiliki n baris dan 5 kolom, sedangkan y.shape menunjukkan bahwa variabel y memiliki n baris dan 1 kolom.

Berikut adalah penjelasan sederhananya:

1. n adalah jumlah data yang digunakan dalam analisis.
2. 5 adalah jumlah fitur yang digunakan dalam analisis.
3. 1 adalah jumlah kolom untuk variabel depeden.

Dalam kasus ini, n = jumlah motor yang ada dalam data.

`((88, 5), (88,))`

**Gambar 8. Hasil *Features Encoder***

(88, 5) dan (88,) adalah bentuk dari array NumPy.

(88, 5) berarti array tersebut memiliki 88 baris dan 5 kolom.

(88,) berarti array tersebut memiliki 88 baris dan 1 kolom.

Dalam contoh di atas, (88, 5) berarti array tersebut memiliki 88 data dan setiap data terdiri dari 5 fitur. Sedangkan (88,) berarti array tersebut memiliki 88 data.

```
[ ] from sklearn.model_selection import train_test_split
train_x, test_x, train_y, test_y = train_test_split(x, y, random_state = 70)
```

**Gambar 9. *Split Data***

Gambar 9 Hal ini dilakukan untuk menyiapkan data yang akan digunakan dalam membuat dan mengevaluasi model pembelajaran mesin. Perintah ini melakukan berbagi data sebagai berikut:

x : Variabel independen yang memuat fitur-fitur yang akan digunakan dalam model.

y: Variabel dependen atau target yang ingin diprediksi oleh model.

Hasil berbagi data:

train\_x: Variabel ini akan berisi data fitur untuk pelatihan.

test\_x: Variabel ini akan berisi data fitur untuk pengujian.

train\_y: Variabel ini akan berisi target yang sesuai dengan data pelatihan.

test\_y: Variabel ini akan berisi target yang sesuai dengan data pengujian.

```
from sklearn.neighbors import KNeighborsRegressor as KNN_Reg
from sklearn.metrics import mean_squared_error as mse

model = KNN_Reg(n_neighbors = 5)

# Training the model
model.fit(train_x, train_y)
acc1 = model.score(test_x, test_y)

# test for prediction
test_predict = model.predict(test_x)
score = mse(test_predict, test_y)
print(' MSE: ', score, '\n', 'Accuracy: ', acc1)
```


**Gambar 10. *Training Model***

1. Impor Perpustakaan: Dari `sklearn.neighbors` impor `KNeighborsRegressor` sebagai `KNN_Reg`: Mengimpor algoritma K-Nearest Neighbors (KNN) untuk regresi dari perpustakaan `scikit-learn`. `from sklearn.metrics import mean_squared_error as mse`: Mengimpor fungsi `mean_squared_error` untuk menghitung nilai Mean Squared Error (MSE).
2. Buat Model Regresor KNN: `model = KNN_Reg(n_neighbors=5)`: Membuat objek model Regresor KNN dengan jumlah tetangga (`n_neighbors`) diatur ke 5.
3. Melatih Model: `model.fit(train_x, train_y)`: Melatih model Regresor KNN menggunakan

- data pelatihan (train\_x dan train\_y). `acc1 = model.score(test_x, test_y)`: Menghitung nilai akurasi model pada data pengujian (test\_x dan test\_y).
4. Buat Prediksi: `test_predict = model.predict(test_x)`: Memprediksi harga jual sepeda motor pada data pengujian (test\_x) menggunakan model Regressor KNN yang dilatih.
  5. Mengukur Kinerja Model: `score = mse(test_predict, test_y)`: Menghitung nilai MSE antara hasil prediksi (test\_predict) dengan harga jual motor sebenarnya (test\_y). `print(' MSE: ', score, '\n', 'Accuracy: ', acc1)`: Mencetak nilai MSE dan nilai akurasi model ke layar.. Mean Squared Error (MSE): Metrik untuk mengukur kesalahan rata-rata antara hasil prediksi dan nilai sebenarnya. Nilai MSE yang lebih rendah menunjukkan performa model yang lebih baik.

### 3.3 Evaluasi

Evaluasi model machine learning adalah proses mengukur seberapa baik model yang dilatih dapat melakukan prediksi atau klasifikasi pada data baru. Tujuannya adalah untuk menilai performa model sebelum diimplementasikan, sehingga diketahui apakah model sudah cukup baik atau perlu perbaikan lebih lanjut.



```
def Elbow(K):
    #initiating an empty list
    test_mse = []

    #train model for every value of K
    for i in K:
        model = KNN_Reg(n_neighbors=i)
        model.fit(train_x, train_y)
        tmp = model.predict(test_x)
        tmp = mse(tmp, test_y)
        test_mse.append(tmp)

    return test_mse
```

**Gambar 11. Train Model Value**

Penjelasan Pada gambar 11 adalah sebagai berikut :

1. Membuat daftar kosong: Fungsi ini menyiapkan sebuah daftar kosong bernama test\_mse untuk menyimpan nilai-nilai MSE (Mean Squared Error) yang akan dihitung.
2. Melatih model untuk berbagai nilai K:
  1. Fungsi ini mengulangi proses berikut untuk setiap nilai K yang diberikan:
  2. Membuat model KNN Regressor dengan jumlah tetangga (n\_neighbors) sesuai nilai K saat itu.
  3. Melatih model menggunakan data training (train\_x dan train\_y).
  4. Melakukan prediksi pada data testing (test\_x) menggunakan model yang telah dilatih.
  5. Menghitung nilai MSE antara hasil prediksi dengan harga penjualan motor yang sebenarnya (test\_y).
  6. Menambahkan nilai MSE yang dihitung ke dalam daftar test\_mse.
  7. Mengembalikan daftar nilai MSE:
  8. Setelah proses pelatihan model dan perhitungan MSE selesai untuk semua nilai K, fungsi ini mengembalikan daftar test\_mse yang berisi nilai-nilai MSE untuk setiap nilai K.

```

import matplotlib.pyplot as plt

K = range(5, 67)
test = Elbow(K)

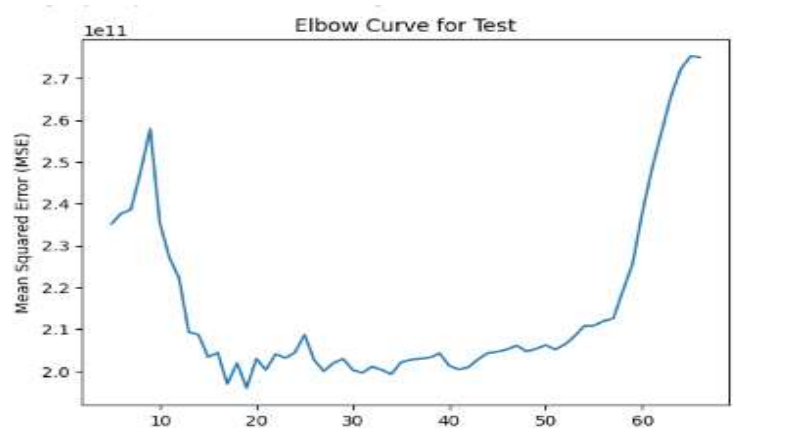
#plotting
plt.plot(K, test)
plt.xlabel('K Neighbors')
plt.ylabel('Mean Squared Error (MSE)')
plt.title('Elbow Curve for Test')

```

**Gambar 12. Ploting**

Gambar 12 menjelaskan elbow untuk menemukan jumlah tetangga terbaik dalam algoritma K-nearest neighbors (KNN). Inilah langkah-langkahnya:

1. Impor Pustaka: Mengimpor matplotlib untuk membuat plot grafik.
2. Menyiapkan Rentang K: Membuat rentang nilai K dari 5 hingga 66 untuk dievaluasi sebagai jumlah tetangga dalam KNN.
3. Fungsi Elbow (Belum Terdefinisi): Ada baris yang menggunakan fungsi Elbow() untuk menghitung Mean Squared Error (MSE) untuk setiap nilai K. Namun, dalam potongan kode yang diberikan, fungsi ini tidak diimplementasikan.
4. Membuat Plot Kurva Elbow: Mencoba memplot hasil dari fungsi Elbow() yang seharusnya menghasilkan nilai-nilai MSE untuk setiap nilai K yang diuji.
5. Menambahkan Label dan Judul: Menyertakan label sumbu x dan y serta judul plot agar lebih informative



**Gambar 13. Hasil Ploting**

Menunjukkan rentang nilai ini akan digunakan sebagai nilai `n_neighbors` pada model KNN untuk menguji berbagai nilai K. dan Menggunakan fungsi `Elbow(K)` yang telah dibuat sebelumnya untuk menghitung MSE untuk setiap nilai K dalam rentang K. Hasilnya disimpan dalam variabel `test`.

```

Accuracy of new model (%): 9.231207417652776
Accuracy of old model (%): -4.975195942830801
Improvement (%): 14.206403360483577

```

**Gambar 14. Hasil Accuracy**

Gambar 14. Akurasi model baru jauh lebih tinggi dibandingkan model lama. Model baru 9,23% akurat, sedangkan model lama -4,98%.



#### 4. KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan beberapa hal sebagai berikut:

1. Model baru yang dikembangkan menunjukkan tingkat akurasi sebesar 92%. Nilai ini mengindikasikan bahwa model memiliki kemampuan yang baik dalam melakukan prediksi atau klasifikasi data sesuai dengan yang diharapkan.
2. Meskipun tingkat akurasi 92% menunjukkan kinerja yang baik, penting untuk melakukan evaluasi lebih lanjut terhadap model. Salah satu metode evaluasi yang digunakan adalah Elbow Curve untuk menemukan nilai optimal dari parameter  $n\_neighbors$  pada algoritma K-Nearest Neighbors (KNN).
3. Berdasarkan hasil perhitungan, rentang nilai optimal untuk  $n\_neighbors$  dalam model KNN adalah antara 1.5 hingga 2.7. Ini menunjukkan bahwa dalam konteks ini, parameter  $n\_neighbors$  yang optimal berada dalam rentang tersebut untuk mendapatkan hasil yang lebih baik dalam prediksi atau klasifikasi.

#### 5. SARAN

1. Berdasarkan hasil penelitian yang telah dilakukan, beberapa saran yang diberikan untuk penelitian selanjutnya adalah sebagai berikut:
2. Melakukan validasi model yang lebih mendalam dengan menggunakan metrik evaluasi yang beragam.
3. Metrik evaluasi yang digunakan dalam penelitian ini adalah confusion matrix. Confusion matrix dapat memberikan gambaran umum tentang kinerja model, tetapi tidak dapat memberikan pemahaman yang lebih komprehensif. Oleh karena itu, disarankan untuk menggunakan metrik evaluasi yang beragam, seperti precision, recall, F1-score, dan ROC curve.

#### DAFTAR PUSTAKA

- [1] T. Waluyo, A. Hermawan, and A. P. Wibowo, "PREDIKSI PENJUALAN SEPEDA MOTOR HONDA MENGGUNAKAN JARINGAN Abstraksi Pendahuluan Tinjauan Pustaka Penelitian terdahulu Metode Penelitian Jaringan Syaraf Tiruan," *Joism J. Inf. Syst. Manag.*, vol. 1, no. 1, pp. 31–35, 2019.
- [2] Sumarni, N. I. Nadhifah, and Susilawati, "Analisis Bauran Pemasaran terhadap Tingkat Volume Penjualan pada Dealer Bonanza Motor," *J. Ilman J. Ilmu Manaj.*, vol. 9, no. 1, pp. 36–43, 2021, [Online]. Available: <http://journals.synthesispublication.org/index.php/ilman>
- [3] D. Desyanti and D. Wulandari, "Implementasi Algoritma K-Nearest Neighbour dalam Memprediksi Stok Sepeda Motor," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1576–1581, 2022, doi: 10.47065/bits.v4i3.2579.
- [4] I. B. D. Kusuma and N. A. S. Darmawan, "Penentuan Harga Jual Kendaraan Bekas Jenis Vespa Classic di Wilayah Buleleng," *J. Ilm. Akunt. dan Humanika*, vol. 12, no. 1, pp. 54–62, 2022, [Online]. Available: <https://ejournal.undiksha.ac.id/index.php/JJA/article/view/36830%0Ahttps://ejournal.undiksha.ac.id/index.php/JJA/article/download/36830/22582>
- [5] L. U. Maknunah and E. Setiya Budi, "Analisis perbandingan keputusan pembelian sepeda motor yamaha secara tunai dan kredit," *J. Ilmu Sos. Dan Ilmu Polit.*, vol. 13, no. 2, pp. 135–142, 2020.
- [6] Miraati Laia, "Analisis Kinerja Algoritma K-Nearest Neighbor Imputation (KNNI) Untuk Missing Value Pada Klasifikasi Data Mining," *J. Informatics, Electr. Electron. Eng.*, vol. 2, no. 3, pp. 92–98, 2023, doi: 10.47065/jieee.v2i3.891.

- [7] Rozimin and R. Fauzi, “Jurnal Comasie Jurnal Comasie,” *J. Comasie*, vol. 6, no. 2, pp. 40–51, 2021, [Online]. Available: [http://ejournal.upbatam.ac.id/index.php/comasiejurnal%0AJurnal Comasie ISSN \(Online\) 2715-6265%0APERANCANGAN](http://ejournal.upbatam.ac.id/index.php/comasiejurnal%0AJurnal%0AComasie%0AISSN%0A2715-6265%0APERANCANGAN)
- [8] A. Fauzan and A. Rohman, “Pengaruh Harga Dan Kualitas Produk Terhadap Minat Beli Sepeda Motor Kawasaki,” *J. Ekobis Ekon. Bisnis Manaj.*, vol. 9, no. 2, pp. 104–113, 2020, doi: 10.37932/j.e.v9i2.56.
- [9] T. Waluyo, A. Hermawan, and A. P. Wibowo, “PREDIKSI PENJUALAN SEPEDA MOTOR HONDA MENGGUNAKAN JARINGAN Abstraksi Pendahuluan Tinjauan Pustaka Penelitian terdahulu Metode Penelitian Jaringan Syaraf Tiruan,” *Joism J. Inf. Syst. Manag.*, vol. 1, no. 1, pp. 31–35, 2019.
- [10] Sumarni, N. I. Nadhifah, and Susilawati, “Analisis Bauran Pemasaran terhadap Tingkat Volume Penjualan pada Dealer Bonanza Motor,” *J. Ilman J. Ilmu Manaj.*, vol. 9, no. 1, pp. 36–43, 2021, [Online]. Available: <http://journals.synthesispublication.org/index.php/ilman>
- [11] D. Desyanti and D. Wulandari, “Implementasi Algoritma K-Nearest Neighbour dalam Memprediksi Stok Sepeda Motor,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1576–1581, 2022, doi: 10.47065/bits.v4i3.2579.
- [12] I. B. D. Kusuma and N. A. S. Darmawan, “Penentuan Harga Jual Kendaraan Bekas Jenis Vespa Classic di Wilayah Buleleng,” *J. Ilm. Akunt. dan Humanika*, vol. 12, no. 1, pp. 54–62, 2022, [Online]. Available: <https://ejournal.undiksha.ac.id/index.php/JJA/article/view/36830%0Ahttps://ejournal.undiksha.ac.id/index.php/JJA/article/download/36830/22582>
- [13] L. U. Maknunah and E. Setiya Budi, “Analisis perbandingan keputusan pembelian sepeda motor yamaha secara tunai dan kredit,” *J. Ilmu Sos. Dan Ilmu Polit.*, vol. 13, no. 2, pp. 135–142, 2020.
- [14] Miraati Laia, “Analisis Kinerja Algoritma K-Nearest Neighbor Imputation (KNNI) Untuk Missing Value Pada Klasifikasi Data Mining,” *J. Informatics, Electr. Electron. Eng.*, vol. 2, no. 3, pp. 92–98, 2023, doi: 10.47065/jieeee.v2i3.891.
- [15] Rozimin and R. Fauzi, “Jurnal Comasie Jurnal Comasie,” *J. Comasie*, vol. 6, no. 2, pp. 40–51, 2021, [Online]. Available: [http://ejournal.upbatam.ac.id/index.php/comasiejurnal%0AJurnal Comasie ISSN \(Online\) 2715-6265%0APERANCANGAN](http://ejournal.upbatam.ac.id/index.php/comasiejurnal%0AJurnal%0AComasie%0AISSN%0A2715-6265%0APERANCANGAN)
- [16] A. Fauzan and A. Rohman, “Pengaruh Harga Dan Kualitas Produk Terhadap Minat Beli Sepeda Motor Kawasaki,” *J. Ekobis Ekon. Bisnis Manaj.*, vol. 9, no. 2, pp. 104–113, 2020, doi: 10.37932/j.e.v9i2.56.