KLASIFIKASI ALGORITMA K- NEAREST NEIGHBORS (KNN) UNTUK SEGMENTASI PASAR MOBIL BEKAS BERDASARKAN MEREK DAN HARGA

Muhammad Rafi Azhar, Agus Iskandar, Ahmad Rifqi
Sistem Informasi, Universitas Nasional
Jl. Sawo Manila No.61 Jakarta, Indonesia
muhammadrafiazhar2021@student.unas.ac.id

ABSTRAK

Penelitian ini berfokus pada segmentasi pasar mobil bekas menggunakan algoritma K-Nearest Neighbors (KNN) untuk memahami preferensi konsumen dalam industri otomotif yang kompetitif. Permasalahan utama yang diidentifikasi adalah implementasi KNN dalam mengklasifikasikan mobil bekas ke dalam kategori grade A, B, dan C. Tujuan penelitian adalah menerapkan algoritma KNN berdasarkan merek dan harga mobil serta mengevaluasi efektivitasnya. Metode yang digunakan mencakup pengumpulan data dari platform jual beli mobil bekas seperti OLX, Ibid, dan Mobil123, dengan variasi nilai K (3, 5, 7, dan 9) serta rasio pembagian data (70:30, 80:20, dan 90:10). Hasil penelitian menunjukkan bahwa nilai K=7 dan rasio data training-to-test 80:20 menghasilkan akurasi tertinggi sebesar 79%, dengan kelas C menunjukkan performa terbaik dengan recall 0,94 dan skor F1 0,84. Temuan ini memberikan kontribusi pada pemahaman segmentasi pasar mobil bekas dan dapat digunakan untuk strategi pemasaran yang lebih terarah di industri otomotif.

Kata kunci : *K-Nearest Neighbors (KNN), Segmentasi Pasar, Mobil Bekas*

1. PENDAHULUAN

Industri otomotif selalu berada dalam kondisi yang berubah-ubah, dengan banyak perusahaan yang berlomba-lomba untuk menarik perhatian pelanggan dengan merilis model-model baru dan mengembangkan teknologi inovatif. Jumlah kendaraan bekas yang dapat digunakan dipengaruhi oleh hal ini karena pembeli kaya menginginkan kendaraan dengan fitur keselamatan dan kemajuan teknologi terkini. Salah satu strategi umum yang digunakan oleh distributor dan pengecer adalah membeli kendaraan bekas dari produsen dan menjualnya kembali ke pelanggan [1].

Dengan semakin banyaknya mobil di pasaran dibandingkan sebelumnya, semakin sulit bagi pembeli untuk menemukan mobil yang sesuai dengan kebutuhan mereka dalam hal fitur seperti girboks, jarak tempuh, jenis bahan bakar, dan tahun produksi. Jika kita ingin mengetahui apa yang diinginkan masyarakat, kita bisa menggunakan K-Nearest Neighbours (KNN) [2].

Metode K-Nearest Neighbors (KNN) merupakan salah satu algoritma pembelajaran mesin yang sederhana namun efektif dalam melakukan klasifikasi dan regresi. KNN bekerja dengan cara mencari k objek terdekat dalam data pelatihan yang paling mirip dengan data baru yang akan diklasifikasikan [3].

Dalam penelitian tentang prediksi harga mobil bekas, penggunaan algoritma K-Nearest Neighbor (KNN) menunjukkan hasil yang menjanjikan. Penelitian ini melibatkan pengumpulan data mobil bekas dari situs Kaggle dan menerapkan model KNN untuk menganalisis harga berdasarkan berbagai faktor seperti tipe bahan bakar, warna, model, dan jarak tempuh [4].

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian terdahulu menunjukkan berbagai pendekatan dalam memanfaatkan algoritma K-Nearest Neighbors (KNN) untuk segmentasi pasar dan prediksi harga mobil bekas. Surya Negara membandingkan KNN dengan algoritma Random Forest dalam memprediksi harga mobil, menemukan bahwa Random Forest lebih unggul dengan akurasi mencapai 96,38% [1].

Naufal mengevaluasi efektivitas KNN dalam membantu pembeli memilih mobil dari situs e-commerce, dengan hasil menunjukkan bahwa metode jarak Euclidean memberikan akurasi tertinggi sebesar 94,40% [2].

Firmansyah menerapkan KNN dalam sistem pendukung keputusan untuk memprediksi penjualan mobil bekas, menghasilkan akurasi 63,89% [5]. Penelitian oleh Samruddhi dan Ashok Kumar juga mengusulkan model KNN yang mencapai akurasi sekitar 85% dalam memprediksi harga mobil bekas berdasarkan berbagai faktor [4].

2.2. Segmentasi Pasar

Segmentasi pasar didefinisikan sebagai "praktik membagi pasar menjadi pasar yang lebih kecil dan lebih mudah dikelola yang ditentukan oleh karakteristik bersama, seperti keinginan, kebutuhan, dan pola perilaku [6].

Di antara pasar yang beragam. Pembeli mobil dapat dikategorikan ke dalam merek premium, kelas menengah, dan ekonomi melalui segmentasi pasar. Analisis dan prediksi harga mobil bekas dan segmentasi pasar telah menjadi subjek penelitian terbaru. Menurut Septian dkk. (2023), telah dibuat sistem pendukung keputusan pemilihan mobil bekas dengan menggunakan algoritma K-Nearest Neighbor

(K-NN). Sistem ini mempertimbangkan variabel termasuk harga, tahun pembuatan, kapasitas penumpang, dan jenis transmisi.[3]

2.3. Merek dan Harga dalam Industri Otomotif

Bisnis mobil sangat dipengaruhi oleh merek dan harga. Penetapan harga yang kompetitif dan persepsi masyarakat yang baik terhadap merek mempengaruhi kecenderungan pelanggan untuk melakukan pembelian. Kombinasi keduanya memberikan pengaruh yang kuat dan menguntungkan terhadap pilihan pembelian konsumen [7].

Sebagai akibat dari meningkatnya pendapatan yang dapat dibelanjakan, jurnal ini melaporkan bahwa sektor otomotif Indonesia sedang berkembang pesat. Hal ini menjadi pertanda baik bagi penjualan mobil di masa depan [8].

2.4. Klasifikasi K-Nearest Neighbor untuk Segmentasi Pasar Mobil bekas

Karena kemampuannya dalam mengelompokkan data berdasarkan karakteristik bersama (misalnya, harga dan merek) dan untuk mengklasifikasikan objek menggunakan informasi tentang lingkungan sekitarnya, algoritma K-Nearest Neighbor (K-NN) telah banyak diterapkan dalam studi segmentasi pasar [5].

Untuk meramalkan kelas objek yang label kelasnya belum diketahui, klasifikasi melibatkan pengembangan model (fungsi) yang mengkarakterisasi dan membedakan kelas atau gagasan data. Menurut Kesuma Dinata dan hasdyna teknik klasifikasi kombinasi berbasis voting adalah pendekatan yang mudah dan serbaguna yang dapat digunakan untuk segala jenis tantangan klasifikasi [9].

2.5. Data Mining

Data mining adalah proses analisis data besar untuk menemukan pola, hubungan, dan wawasan yang tersembunyi. Dalam konteks industri otomotif, data mining dapat digunakan untuk menganalisis data penjualan, perilaku konsumen, dan tren pasar untuk memberikan informasi berharga yang dapat meningkatkan strategi pemasaran dan pengembangan produk. Teknik-teknik dalam data mining, seperti analisis kluster dan pengenalan pola, memungkinkan perusahaan untuk memahami lebih dalam tentang preferensi dan perilaku konsumen.[10]

Dengan demikian, data mining tidak hanya meningkatkan penjualan tetapi juga membantu dalam pengambilan keputusan yang lebih baik, memfasilitasi inovasi produk, dan meningkatkan kepuasan pelanggan. Penggunaan data mining dalam industri otomotif memberikan wawasan yang lebih luas dan mendalam tentang dinamika pasar, yang sangat penting dalam menghadapi tantangan kompetitif saat ini [11].

2.6. Algoritma *K – Nearest Neighbor (KNN)*

Salah satu algoritma yang belajar dari pendahulunya adalah metode K-NN (K Nearest Neighbors), yang digunakan untuk tujuan klasifikasi. Teknik K-NN didasarkan pada gagasan bahwa data pelatihan paling akurat diwakili oleh sampel yang sangat dekat dengan kelas tersebut. Untuk memilih jumlah tetangga terdekat, teknik ini menggunakan nilai k yang ditentukan dengan mengurutkan nilai terendah dari hasil perhitungan jarak. Pemilihan lingkungan dengan suara tertinggi menentukan prosedur kategorisasi. [12].

$$d_{(a,b} = \sqrt{\sum_{g=1}^{p}(x_{ag} - x_{bg})^2} \quad (1)$$

dengan:

$d_{(a,b)} = jarak\ antara\ objek\ a\ dengan\ b$
$x_{ag} = nilai\ objek\ data\ training\ a\ pada\ variabel\ ke - g$
$x_{bg} = nilai\ objek\ data\ testing\ b\ pada\ variabel\ ke - g$
$p = banyaknya\ variabel\ bebas$

2.7. Perhitungan Manual Menggunakan Excel

Excel berfungsi sebagai alat untuk menghitung jarak euclidean secara manual serta mengelola dan menganalisis data. Setelah menghitung jarak, data diurutkan berdasarkan nilai terdekat, dan kategori untuk data baru ditentukan berdasarkan mayoritas dari K tetangga terdekat yang diidentifikasi. sehingga memberikan pemahaman yang lebih mendalam tentang efektivitas algoritma K-NN dalam klasifikasi penerima dana bantuan [13].

2.8. Python

Untuk menggunakan metode K-Nearest Neighbor (K-NN) untuk klasifikasi kualitas apel, bahasa pemrograman Python sangat penting. Python dipilih karena kemudahan penggunaan dan penyertaan beberapa perpustakaan yang menyederhanakan proses penambangan data dan pembelajaran mesin. Proses preprocessing, pemisahan data training dan testing, serta evaluasi kinerja model dilakukan dengan memanfaatkan kemampuan Python untuk menangani dataset yang besar [14].

2.9. Confusion Matrix

Matriks konfusi merupakan bagian integral dari analisis klasifikasi, menampilkan jumlah TP, TN, FP, dan FN. Keakuratan model ditentukan menggunakan matriks ini, dengan menghitung tingkat positif sebenarnya (TPR) dan negatif sebenarnya (TNR). Penelitian menunjukkan bahwa TPR dan TNR dapat dipahami melalui hubungan geometris yang diwakili oleh sudut-sudut pada segitiga siku-siku dari komponen matriks. Dengan menganalisis matriks, peneliti dapat menentukan ambang optimal untuk meningkatkan akurasi klasifikasi dan menilai daya diskriminasi model, di mana jarak antara titik A dan B dalam plot TPR-TNR menjadi indikator performa klasifikasi. [15].

2.10. Rapid Miner

Perangkat lunak pengolah data RapidMiner memberikan jawaban atas permasalahan dalam data mining, text mining, dan analisis prediktif. Operator yang berperan memanipulasi data memudahkan pengguna dalam menghitung data dalam jumlah besar di aplikasi ini. Node operator menghubungkan data, dan node hasil menampilkan hasil akhir; grafik kemudian digunakan untuk menyajikan hasilnya [11].

3. METODE PENELITIAN

Dalam desain penelitian ini, yang menggunakan pendekatan kuantitatif. Pendekatan kuantitatif dipilih karena data penelitian ini berbentuk numerik, serta untuk memberikan hasil analisis yang objektif dan terukur.

3.1. Lokasi Penelitian

Penelitian ini dilaksanakan tanpa melibatkan proses pengumpulan data di lapangan, melainkan dengan mengambil data sekunder dari sebuah situs dataset jual beli mobil bekas. Situs tersebut, yang menyediakan dataset kendaraan bekas, sangat relevan untuk tujuan segmentasi berdasarkan merek dan harga.

3.2. Waktu Penelitian

Selama bulan Oktober sampai Februari 2025, penelitian ini dilakukan.

3.3. Penentuan Subjek

Subjek penelitian ini terdiri dari data mobil bekas yang dijual di pasar. Fokus utama penelitian terletak pada data yang memberikan informasi mengenai merek dan harga mobil bekas. Kedua variabel ini dipilih sebagai faktor kunci dalam proses segmentasi pasar, mengingat preferensi konsumen sering kali bergantung pada merek tertentu serta rentang harga yang sesuai dengan anggaran mereka

3.4. Fokus Penelitian

Penelitian ini berfokus pada beberapa aspek berikut:

a. Implementasi algoritma K-Nearest Neighbors (KNN): Penelitian ini menggunakan Rapidminer sebagai alat untuk mengimplementasikan algoritma K-Nearest Neighbors (KNN) untuk mengklasifikasikan segmen pasar mobil bekas dan menggunakan bahasa pemrograman Python.
b. Penilaian Efektivitas KNN dalam Segmentasi Pasar: Penelitian ini mengevaluasi keefektifan algoritma KNN dalam membagi mobil bekas ke dalam segmen pasar yang berbeda berdasarkan merek dan harga.
c. Evaluasi Performa Algoritma KNN: Selain mengimplementasikan, penelitian ini juga mengukur performa algoritma KNN yang digunakan dalam Rapidminer untuk mengetahui sejauh mana KNN mampu melakukan segmentasi pasar dengan akurasi tinggi.

3.5. Sumber Data

Penelitian ini menggunakan data mobil bekas yang diperoleh dari dua platform penjualan online utama, yaitu OLX, Ibid, dan Mobil123, yang diperoleh mencakup informasi-informasi penting seperti:

a. Merek Mobil: Variabel ini dipakai untuk mengidentifikasi preferensi konsumen terhadap merek tertentu.
b. Harga Mobil: Informasi mengenai harga sangatlah penting dalam klasifikasi segmen pasar.
c. Grade Mobil Bekas: Grade mobil bekas dikategorikan menjadi A, B, dan C. Kategori ini memberikan gambaran tentang kondisi fisik dan mekanis mobil.

OLX, Ibid, dan Mobil123 dipilih sebagai sumber data karena menyediakan variabel-variabel yang sesuai.

3.6. Teknik Pengumpulan Data

Berikut tahapan pendekatan pengumpulan data yang digunakan dalam penelitian ini:

a. Informasi yang Diperoleh dari OLX, Ibid, dan Mobil123: Informasi yang digunakan untuk penelitian ini diambil dari situs web OLX, Ibid, dan Mobil123
b. Pengumpulan data selanjutnya, tahap selanjutnya adalah validasi data, yang meliputi pemeriksaan keakuratan dan kelengkapan data.

3.7. Desain Penelitian

Python berfungsi sebagai bahasa utama untuk analisis data dalam desain penelitian ini, yang menggunakan pendekatan kuantitatif.. Pendekatan kuantitatif dipilih karena data penelitian ini berbentuk numerik, serta untuk memberikan hasil analisis yang objektif dan terukur. Algoritma K-Nearest Neighbors (KNN) digunakan dalam mengklasifikasikan mobil bekas ke dalam segmen pasar berdasarkan data numerik (harga) dan kategorikal (merek mobil).

3.8. Tahap Penelitian

Beberapa tahapan menjadi kerangka penelitian dalam penelitian ini. Gambar 3.1 memberikan rincian lengkap tahapan penelitian.

Gambar 1. Tahapan penelitian

3.8.1. Identifikasi Masalah

Pada tahap ini, peneliti mengidentifikasi dan mendefinisikan masalah yang ingin diselesaikan. Ini mencakup pemahaman tentang apa yang menjadi fokus penelitian.

3.8.2. Studi Literatur

Menemukan materi terkait dan penelitian sebelumnya merupakan langkah awal dalam melakukan tinjauan pustaka, yang dilanjutkan dengan identifikasi masalah. Latar belakang dan landasan teori saat ini mungkin dapat lebih dipahami dengan informasi ini.

3.8.3. Pengumpulan Data

Dalam proses penelitian ini pengumpulan data merupakan tahap pertama. Informasi yang dikumpulkan bersumber dari berbagai situs jejaring sosial, termasuk OLX, Ibid, dan Mobil123.

3.8.4. Preprocessing Data

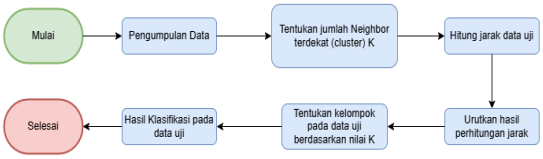
Tahap di mana data dibersihkan dan dipersiapkan untuk analisis. Ini mencakup penghapusan data yang hilang, normalisasi, encoding kategori, dan transformasi lain yang diperlukan agar data siap digunakan dalam model.

3.8.5. Klasifikasi Algoritma KNN

Penelitian dimulai dengan pengumpulan data dan penentuan jumlah tetangga terdekat (K) sebagai parameter utama algoritma KNN. Setelah itu, jarak antara data uji dan data pelatihan dihitung dan diurutkan untuk menentukan tetangga terdekat. Berdasarkan mayoritas kelas dari K tetangga, data uji diklasifikasikan ke dalam kelompok tertentu. Hasil klasifikasi ini dievaluasi untuk menilai performa algoritma, yang kemudian dianalisis dan di dokumentasikan.

3.8.6. Hasil Pengujian Algoritma KNN

Dalam penelitian ini, matriks konfusi digunakan untuk memahami dan menilai kemanjuran teknik kategorisasi. Pengujian matriks konfusi dijelaskan pada Tabel 1 di bawah ini.



Gambar 2. Klasifikasi algoritma KNN

Tabel 1. Hasil algoritma KNN

		Actual Data	
		Actual Positive	***Actual Negative***
Predicted Data	**Predicted Positive**	True Positive (TP)	False Positive (FP)
	Predicted Negative	False Negative (FN)	True Negative (TN)

Keterangan:

a. Data yang benar-benar kelas (+) diidentifikasi sebagai kelas benar (+) dalam situasi true positif (TP).
b. FN, atau negatif palsu, mengacu pada data yang seharusnya diklasifikasikan sebagai kelas (+) namun malah diidentifikasi sebagai kelas (-).
c. FP, atau positif palsu, mengacu pada data yang seharusnya diklasifikasikan sebagai kelas (-), namun malah diidentifikasi sebagai kelas (+).
d. Data yang benar-benar berkelas (-) diidentifikasi sebagai kelas benar (-), fenomena yang disebut true negative (TN).

3.8.7. Kesimpulan dan Saran

Langkah terakhir adalah peneliti menarik kesimpulan dan memberikan rekomendasi untuk studi lebih lanjut atau penggunaan praktis informasi tersebut.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Dalam proses penelitian ini pengumpulan data merupakan tahap pertama. Informasi yang dikumpulkan bersumber dari berbagai situs jejaring sosial, termasuk OLX, Ibid, dan Mobil123.

Tabel 2. Pengumpulan data

No	Merek	Harga	Grade
1	Ford Everest Trend Standard (2010)	93100000	C
2	Honda Brio Satya Standard (2019)	172100000	A
…	…	…	…
499	Chevrolet Captiva Ultimate (2019)	196600000	A
500	Nissan Livina E (2016)	133100000	C

Tabel 2 di atas menunjukkan klasifikasi data yang dibuat menggunakan 500 mobil bekas yang terdaftar di OLX, Ibid, dan Mobil123. Mobil-mobil tersebut diurutkan berdasarkan merek, harga, dan grade.

4.2. Preprocessing data

```
[147] import pandas as pd
      from sklearn.model_selection import train_test_split
      from sklearn.preprocessing import LabelEncoder
      from sklearn.metrics import classification_report, confusion_matrix, ConfusionMatrixDisplay
      import matplotlib.pyplot as plt
      import numpy as np
```

Gambar 3. Preprocessing data

Kode di atas mempersiapkan berbagai pustaka yang diperlukan untuk analisis data, pembelajaran mesin, dan visualisasi hasil model

```
# Membaca dataset
data = pd.read_csv('optimized_grade_DATA_500.csv', sep=';')
```

Gambar 4. Membaca dataset

Kode ini bertujuan untuk membaca dataset dari file CSV yang dipisahkan oleh titik koma dan menyimpannya dalam variabel data

```
# Mengubah label kategori menjadi numerik
label_encoder = LabelEncoder()
data['Grade'] = label_encoder.fit_transform(data['Grade'])
```

Gambar 5. Mengubah label

Mengubah label grade (misalnya 'A', 'B', dan 'C') diubah menjadi format numerik menggunakan LabelEncoder.

```
# Membersihkan kolom 'harga' untuk menghapus titik dan mengonversi ke numerik
data['harga'] = data['harga'].str.replace('.', '', regex=False).astype(float)
```

Gambar 6. Membersihkan data

Membersihkan kolom harga Menghapus titik dari kolom 'harga' dan mengonversinya menjadi tipe data numerik.

```
# Memisahkan fitur dan target
X = data[['Harga (IDR)']]
y = data['Grade']
```

Gambar 7. Memisahkan fitur

Kode ini bertujuan untuk memisahkan fitur (dalam hal ini, kolom "Harga (IDR)" dan target (kolom "Grade") dari dataset.

```
# Membagi dataset
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Gambar 8. Membagi dataset

Dengan kode ini, kumpulan data dipartisi menjadi dua bagian: pelatihan dan pengujian. Model dilatih menggunakan 80% data, sedangkan 20% sisanya digunakan untuk pengujian.

4.3. Perhitungan manual Algoritma K-Nearest Neighbor

Berikut ini merupakan representasi perhitungan manual sederhana dari algoritma K-Nearest Neighbor, Pada Tabel 3 merupakan sampel data mobil bekas :

Tabel 3. Data latih

Merek	Harga	Grade
Ford Everest Trend Standard (2010)	93,100,000	C
Honda Brio Satya Standard (2019)	172,100,000	A
Hyundai Creta Trend Standard (2011)	87,900,000	C
Mitsubishi Outlander Sport Standard (2021)	290,500,000	A
Chevrolet Spin LTZ Standard (2022)	237,700,000	B

Berikut ini merupakan representasi perhitungan manual sederhana dari algoritma K-Nearest Neighbor, Pada Tabel 4 berikut merupakan sampel data uji mobil bekas :

Tabel 4. Data uji

Merek	Harga	Grade
Nissan Livina E (2016)	133,100,000	?

a. Rumus Euclidean Distance:

$$d = \sqrt{(x_2 - x_1)^2} \quad (2)$$

Hitung Jarak

Ford Everest Trend Standard (2010):

$$d = \sqrt{(93{,}100{,}000 - 133{,}100{,}000)^2} = \sqrt{(-40{,}000{,}000)^2} = 40{,}000{,}000 \quad (3)$$

Honda Brio Satya Standard (2019):

$$d = \sqrt{(172{,}100{,}000 - 133{,}100{,}000)^2} = \sqrt{(39{,}000{,}000)^2} = 39{,}000{,}000 \quad (4)$$

Hyundai Creta Trend Standard (2011):

$$d = \sqrt{(87{,}900{,}000 - 133{,}100{,}000)^2} = \sqrt{(-45{,}200{,}000)^2} = 45{,}200{,}000 \quad (5)$$

Mitsubishi Outlander Sport Standard (2021):

$$d = \sqrt{(290{,}500{,}000 - 133{,}100{,}000)^2} = \sqrt{(157{,}400{,}000)^2} = 157{,}400{,}000 \quad (6)$$

Chevrolet Spin LTZ Standard (2022):

$$d = \sqrt{(239{,}700{,}000 - 133{,}100{,}000)^2} = \sqrt{(104{,}600{,}000)^2} = 104{,}600{,}000 \quad (7)$$

b. Mengurutkan Jarak

Dengan K = 3, kita ambil 3 mobil terdekat berdasarkan jarak:

- Honda Brio Satya Standard (2019): 39,000,000
- Ford Everest Trend Standard (2010): 40,000,000
- Hyundai Creta Trend Standard (2011): 45,200,000
- Chevrolet Spin LTZ Standard (2022): 104,600,000
- Mitsubishi Outlander Sport Standard (2021): 157,400,000

c. Menentukan Grade

Kita ambil 3 tetangga terdekat:

- Honda Brio Satya Standard (A)
- Ford Everest Trend Standard (C)
- Hyundai Creta Trend Standard (C)

d. Voting

- A: 1
- C: 2

e. Kategori Terdekat

Berdasarkan algoritma KNN dengan K = 3, model memprediksi bahwa Nissan Livina E (2016) termasuk dalam kelas C.

4.4. Klasifikasi Algoritma KNN

Menemukan nilai K memungkinkan dilakukannya perhitungan klasifikasi K-NN. Nilai K yang digunakan dalam penyelidikan ini adalah 3, 5, 7, dan 9.

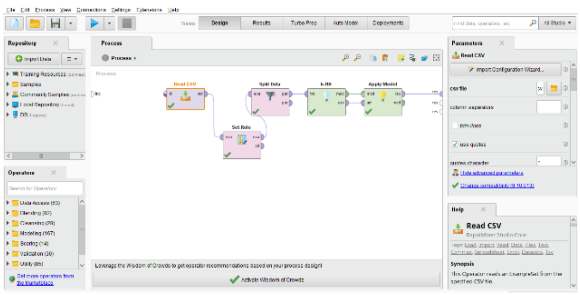
```
from sklearn.neighbors import KNeighborsClassifier
# Membuat model KNN
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(X_train, y_train)
```

Gambar 9. Klasifikasi algoritma KNN

Kode ini digunakan untuk membuat dan melatih model K-Nearest Neighbors (KNN) menggunakan data latih yang telah disiapkan. Model ini akan digunakan untuk melakukan klasifikasi berdasarkan data baru dengan mempertimbangkan 5 tetangga terdekat.

4.5. Klasifikasi Proses Penerapan Algoritma KNN

Pada langkah ini kita melihat bagaimana mengklasifikasikan mobil bekas berdasarkan harga dan spek menggunakan algoritma K-Nearest Neighbors (KNN) RapidMiner.



Gambar 10. Klasifikasi proses

Dataset dibaca dengan Read CSV, lalu atribut target ditentukan melalui Set Role. Data kemudian dibagi menggunakan Split Data untuk pelatihan dan pengujian. Algoritma K-NN diterapkan, dan hasilnya digunakan pada Apply Model untuk memprediksi kategori kendaraan.

4.6. Hasil Pengujian

Hasil pengujian ini bertujuan untuk mengevaluasi kinerja model yang dikembangkan dalam pengklasifikasian data.

Row No.	Grade	prediction(G...	confidence(A)	confidence(B)	confidence(C)	Merek	Harga (IDR)
1	C	C	0	0	1	Ford Everest ...	93100000
2	A	A	1	0	0	Honda Brio S...	172100000
3	A	A	1	0	0	Mitsubishi Ou...	290500000
4	B	B	0.429	0.571	0	Chevrolet Spi...	237700000
5	B	B	0.333	0.667	0	Daihatsu Sigr...	196800000
6	B	A	0.500	0.500	0	Nissan Sere...	246800000
7	C	C	0	0	1	Ford Everest ...	89000000
8	A	A	1	0	0	Mitsubishi Xp...	329200000
9	A	A	0.500	0.500	0	Mitsubishi Xp...	231300000
10	C	C	0	0	1	Hyundai Tucs...	103700000
11	C	C	0	0	1	Suzuki Balen...	136400000
12	C	C	0	0.417	0.583	Hyundai Tucs...	119200000
13	C	C	0	0.429	0.571	Mazda BT-50 ...	242000000
14	B	B	0	1	0	Toyota Avanz...	155300000
15	B	B	0	0.757	0.243	Mitsubishi Pa...	239500000

Gambar 11. Hasil pengujian

Gambar diatas menunjukkan hasil prediksi klasifikasi mobil bekas menggunakan algoritma K-Nearest Neighbors (KNN) di RapidMiner.

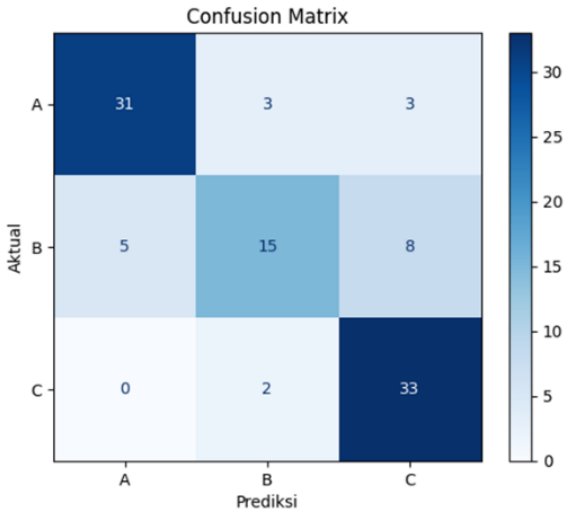
4.6.1. Pengujian Confusion Matrix

Berbagai tahapan pengujian untuk menentukan dan menilai nilai skor f1, presisi, recall, dan akurasi. Penelitian ini menggunakan matriks konfusi untuk pengujian, dengan K = 3,5, 7, dan 9 untuk jumlah tetangga terdekat, serta 70:30, 80:20, dan 90:10 untuk perbandingan. Skor akurasi terbaik yang dicapai pada percobaan sebesar 79,00% dengan menggunakan rasio data training-to-test 80:20 dengan nilai K=7.

Tabel 5. Pengujian confusion matrix

Nilai K	Accuracy		
	70:30	**80:20**	**90:10**
3	71.33%	72.00%	68.00%
5	74.00%	75.00%	72.00%
7	74.00%	**79.00%**	74.00%
9	73.33%	78.00%	78.00%

Berdasarkan temuan yang ditunjukkan pada tabel konfigurasi pengujian yang paling berhasil mencapai akurasi 79,00% menggunakan rasio data pelatihan/pengujian 80:20 dan nilai K=7.



Gambar 12. Confusion matrix

Gambar ini menunjukkan Confusion Matrix model KNN, yang membandingkan kelas aktual dengan prediksi. Model bekerja baik untuk kelas A dan C, dengan 31/37 dan 33/35 data diklasifikasi benar. Namun, kelas B memiliki banyak kesalahan, dengan hanya 15/28 data diklasifikasi benar, sementara 5 salah ke A dan 8 ke C. Berikut ada perhitungan untuk menghitung akurasinya:

$$\text{Akurasi} = \frac{jumlah\ prediksi\ benar}{total\ data}$$

$$\text{Akurasi} = \frac{31+15+33}{37+28+25} = \frac{79}{100} = 0.79 = 79\% \qquad (8)$$

1) Jumlah Prediksi Benar:

- Benar untuk kelas A: 31
- Benar untuk kelas B: 15
- Benar untuk kelas C: 33

Sehingga, jumlah prediksi benar adalah:
31+15+33=79

2) Total Jumlah Prediksi:

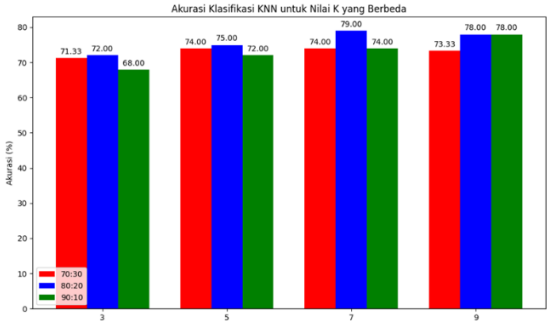
- Total untuk kelas A: 37
- Total untuk kelas B: 28
- Total untuk kelas C: 25

Sehingga, total jumlah prediksi adalah:
37+28+25=100

Dengan hasil $0.79x100 = \mathbf{79}\%$ (9)

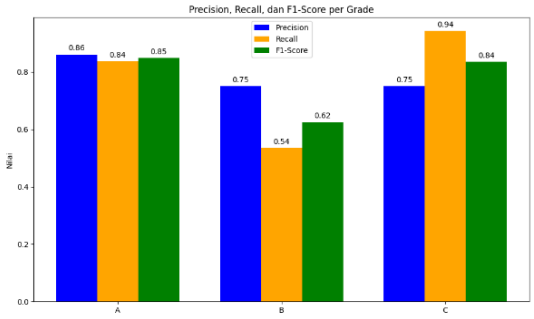
4.6.2. Visualisasi Akurasi Model KNN

Bagian ini menyajikan visualisasi akurasi model K-Nearest Neighbors (KNN) yang telah diuji dengan berbagai nilai K dan pembagian dataset. Grafik yang ditampilkan menunjukkan hubungan antara nilai K dan akurasi yang diperoleh, memberikan gambaran yang jelas tentang performa model.



Gambar 13. Diagram batang

Berdasarkan hasil ini, K = 7 dengan rasio 80:20 adalah konfigurasi terbaik untuk klasifikasi segmentasi pasar mobil bekas.



Gambar 14. Pengujian 80:20 K-7

Diagram menggambarkan Pengujian dengan rasio 80:20 dengan nilai K= 7. Nilai Precision, Recall, dan F1-Score untuk tiga kelas (A, B, dan C) dalam analisis klasifikasi. Berikut ada perhitungan untuk menghitung akurasinya:

a. Grade A

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{31}{31+5} = \frac{31}{36} = 0.86$$

$$\text{Recall} = \frac{TP}{TP+FP} = \frac{31}{31+6} = \frac{31}{37} = 0.84$$

$$\text{F1 Score} = \frac{2\ x\ precision\ x\ recall}{precision + recall} = \frac{2\ x\ 0.86\ x\ 0.84}{0.86 + 0.84} = \frac{1.44}{1.7} = 0.85 \quad (10)$$

b. Grade B

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{15}{15+5} = \frac{15}{20} = 0.75$$

$$\text{Recall} = \frac{TP}{TP+FP} = \frac{15}{15+13} = \frac{15}{28} = 0.54$$

$$\text{F1 Score} = \frac{2\ x\ precision\ x\ recall}{precision + recall} = \frac{2\ x\ 0.75\ x\ 0.54}{0.75 + 0.54} = \frac{0.81}{1.29} = 0.62 \quad (11)$$

c. Grade C

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{33}{33+11} = \frac{33}{44} = 0.75$$

$$\text{Recall} = \frac{TP}{TP+FP} = \frac{33}{33+2} = \frac{33}{35} = 0.94$$

$$\text{F1 Score} = \frac{2\ x\ precision\ x\ recall}{precision + recall} = 2\ \frac{x\ 0.75\ x\ 0.94}{0.75 + 0.94} = \frac{1.41}{1.69} = 0.84 \quad (12)$$

5. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Nearest Neighbor (KNN) untuk segmentasi pasar mobil bekas berdasarkan merek dan harga, dengan konfigurasi terbaik pada K=7 dan rasio data pelatihan/pengujian 80:20, menghasilkan akurasi maksimum 79%. Hasil menunjukkan kelas A dan C berkinerja baik, sementara kelas B kurang memuaskan. Temuan ini menekankan pentingnya pemilihan parameter yang tepat dalam KNN untuk meningkatkan akurasi klasifikasi.

Untuk meningkatkan akurasi model dalam penelitian selanjutnya, disarankan agar menggunakan dataset yang lebih luas dan beragam, sehingga variasi di pasar mobil bekas dapat lebih dipahami. Peneliti juga sebaiknya mengeksplorasi metode lain seperti Random Forest atau Support Vector Machine (SVM) untuk membandingkan efektivitas algoritma dalam segmentasi pasar. Selain itu, penambahan fitur relevan seperti kondisi fisik mobil, riwayat perawatan, atau ulasan pengguna dapat memberikan konteks lebih kaya dan meningkatkan akurasi klasifikasi, menjadikan hasil penelitian lebih informatif bagi industri otomotif.

DAFTAR PUSTAKA

[1] E. Surya Negara, J. Jenderal Ahmad Yani, K. I. Seberang Ulu, and S. Selatan, “Sulaiman et al, Komparasi Algoritma K-Nearest Neoghbors dan Random Forest …….. 337 Komparasi Algoritma K-Nearest Neighbors dan Random Forest Pada Prediksi Harga Mobil Bekas,” 2023. [Online]. Available: www.cardekho.com.

[2] M. F. Naufal, “Finding The Most Desirable Car Using K-Nearest Neighbor From E-Commerce Websites,” *Jurnal ELTIKOM*, vol. 5, no. 1, pp. 25–31, Mar. 2021, doi: 10.31961/eltikom.v5i1.221.

[3] M. A. Septian, M. Assidiq, and C. R. Sari, “SISTEM PENDUKUNG KEPUTUSAN PEMBELIAN MOBIL BEKAS TERBAIK DENGAN METODE K - NEAREST NEIGHBOR (K-NN),” *Journal Peqguruang: Conference Series*, vol. 5, no. 2, p. 783, Nov. 2023, doi: 10.35329/jp.v5i2.4291.

[4] K. Samruddhi and R. Ashok Kumar, “Used Car Price Prediction using K-Nearest Neighbor Based Model,” *International Journal of Innovative Research in Applied Sciences and Engineering*, vol. 4, no. 2, pp. 629–632, Aug. 2020, doi: 10.29027/ijirase.v4.i2.2020.629-632.

[5] Firmansyah, “IMPLEMENTASI METODE K-NEAREST NEIGHBOR DALAM

PERAMALAN PENJUALAN MOBIL BEKAS DI KOTA BATAM," 2021.

[6] E. Vivaldy *et al.*, "ELFANDO BERSAUDARA SENTOSA IN NORTH MINAHASA," vol. 11, no. 1, pp. 866–872, 2023.

[7] F. A. Kuengo, H. Taan, and D. L. Radji, "Pengaruh Citra Merek Dan Harga Terhadap Keputusan Pembelian Mobil Honda Brio Pada Nengga mobilindo Kota Gorontalo," *JAMBURA*, vol. 5, p. 2022, 2022, [Online]. Available: http://ejurnal.ung.ac.id/index.php/JIMB

[8] N. Indriastuty, "PENGARUH KUALITAS PRODUK, HARGA DAN PROMOSI TERHADAP KEPUTUSAN PEMBELIAN MOBIL MEREK TOYOTA DI KOTA BALIKPAPAN," vol. 12, pp. 2086–1117, 2021, doi: 10.36277/geoekonomi.

[9] R. Kesuma Dinata and N. Hasdyna, "KLASIFIKASI SEKOLAH MENENGAH PERTAMA/SEDERAJAT WILAYAH BIREUEN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS BERBASIS WEB," 2020.

[10] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "ANALISIS PERFORMA METODE K-NEAREST NEIGHBOR UNTUK IDENTIFIKASI JENIS KACA," *ILKOM Jurnal Ilmiah*, vol. 11, no. 3, pp. 269–274, Dec. 2019, doi: 10.33096/ilkom.v11i3.489.269-274.

[11] C. H. P. Panjaitan, L. J. Pangaribuan, and C. I. Cahyadi, "Analisis Metode K-Nearest Neighbor Menggunakan Rapid Miner untuk Sistem Rekomendasi Tempat Wisata Labuan Bajo," *Remik*, vol. 6, no. 3, pp. 534–541, Aug. 2022, doi: 10.33395/remik.v6i3.11701.

[12] S. Ulya, M. Arief Soeleman, F. Budiman, and M. Teknik Informatika, "Optimasi Parameter K Pada Algoritma K-NN Untuk Klasifikasi Prioritas Bantuan Pembangunan Desa Optimization of K Parameters in the K-NN Algorithm for Priority Classification of Village Development Assistance," 2021.

[13] R. L. Hasanah *et al.*, "KLASIFIKASI PENERIMA DANA BANTUAN DESA MENGGUNAKAN METODE KNN (K-NEAREST NEIGHBOR)," *Jurnal TECHNO Nusa Mandiri*, vol. 16, no. 1, p. 1, 2019, [Online]. Available: http://nusamandiri.ac.id/

[14] P. Astuti, U. Nusa Mandiri Jln Jatiwaringin Raya No, K. Cipinang Melayu Kecamatan Makasar, J. Timur, and C. Author, "Klasifikasi Kualitas Buah Apel Dengan Algoritma K-Nearest Neighbor (K-NN) Menggunakan Bahasa Pemrograman Python," 2024. [Online]. Available: https://www.kaggle.com/

[15] I. Habib Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," vol. 8, no. 3, 2023.