

Analisis Faktor Yang Mempengaruhi Penjualan Mobil Berdasarkan Model K-Nearest Neighbour

Kiki Mustaqim^{1)*}, Fatia Amalia Maresti²⁾, Siti Salwa Azzahra³⁾, Salsa Paringga Pangestu Ningsih⁴⁾

^{1,2,3,4)}Universitas Logistik dan Bisnis Internasional, Fakultas Logistik, Teknologi dan Bisnis, Program Studi Sains Data.

Received: 9 March 2025

Accepted: 14 August 2025

Published: 1 December 2025



*salwazahra126@gmail.com

Kata Kunci: K-Nearest Neighbor, Klasifikasi Penjualan, Faktor Penjualan Mobil

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Industri otomotif memiliki peran penting dalam pertumbuhan ekonomi global, termasuk di Indonesia. Persaingan antar produsen semakin ketat, sehingga pemahaman terhadap faktor-faktor yang mempengaruhi keputusan konsumen dalam membeli mobil menjadi sangat penting. Penelitian ini bertujuan untuk menerapkan model klasifikasi menggunakan algoritma K-Nearest Neighbors (KNN) untuk memprediksi kategori penjualan mobil ke dalam tiga kategori yaitu sangat laris, cukup laris, dan kurang laris dan penelitian ini menganalisis faktor-faktor yang paling berpengaruh terhadap tingkat penjualan mobil. Data yang digunakan berasal dari penjualan wholesales mobil pada tahun 2023–2024. Proses analisis dilakukan melalui beberapa tahapan, yaitu seleksi data, preprocessing, transformasi, data mining, dan evaluasi model. Hasil penelitian model KNN yang diterapkan menghasilkan akurasi sebesar 71% dengan nilai K=5, yang menunjukkan performa cukup baik dalam memprediksi. Kategori penjualan mobil dan menunjukkan bahwa variabel PS/HP merupakan faktor utama yang mempengaruhi keputusan pembelian, diikuti oleh kategori mobil, jumlah seater, transmisi, jenis bahan bakar, harga, kapasitas tangki, dan kapasitas mesin (CC). Penelitian ini dapat dimanfaatkan oleh industri otomotif dalam strategi pemasaran dan pengembangan produk agar lebih sesuai dengan preferensi pasar.

PENDAHULUAN

Industri otomotif berperan penting dalam pertumbuhan ekonomi global, termasuk di Indonesia. Penjualan mobil memiliki peran yang sangat penting dalam industri otomotif, karena dapat memberikan gambaran tentang keadaan perekonomian suatu negara[1]. Keputusan konsumen dalam membeli mobil dipengaruhi oleh berbagai faktor tidak hanya harga, tetapi juga kapasitas mesin (CC), jenis bahan bakar, sistem transmisi, dan kapasitas tangki bensin. Algoritma K-Nearest Neighbor (KNN) merupakan salah satu metode klasifikasi yang sederhana namun efektif, karena mengklasifikasikan objek berdasarkan kemiripan atau kedekatan data dengan data latih[2]. Banyaknya faktor yang mempengaruhi penjualan mobil, dibutuhkan K-Nearest Neighbors (KNN) yang bertujuan untuk menerapkan model klasifikasi yang dapat memprediksi pada kategori penjualan mobil berdasarkan berbagai fitur seperti spesifikasi (CC), jenis bahan bakar, sistem transmisi, dan kapasitas tangki bensin.

Hal ini membantu mengidentifikasi apakah mobil termasuk dalam klasifikasi kategori sangat laris, cukup laris atau kurang laris dan analisis ini dapat menunjukkan faktor-faktor yang paling berpengaruh terhadap keputusan pembelian mobil pada tingkat penjualan mobil. Hasil analisis tersebut bisa menjadi acuan bagi produsen dalam merancang dan memproduksi mobil yang sesuai dengan tren dan preferensi pasar konsumen. Data Mining adalah proses mengekstraksi dan mengidentifikasi informasi berguna

dan pengetahuan yang relevan dari database besar menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin. Istilah data mining pada hakikatnya adalah suatu bidang ilmu tujuan utamanya menemukan, mempelajari, atau memperoleh pengetahuan dari data dan informasi yang kita miliki [3].

Algoritma K-Nearest Neighbor merupakan bagian dari algoritma supervised learning, yakni hasil dari query instance yang baru, diklasifikasikan berdasarkan data terbanyak dari kategori yang ada. Kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi [4]. Penelitian bertujuan menerapkan model klasifikasi yang dapat memprediksi kategori penjualan mobil serta menganalisis pentingnya setiap faktor dalam menentukan tingkat penjualan mobil. Hasil penelitian diharapkan dapat memberikan wawasan bagi industri otomotif dalam memahami preferensi konsumen serta mendukung pengambilan keputusan terkait strategi pemasaran, penentuan harga, dan pengembangan produk yang sesuai dengan permintaan pasar.

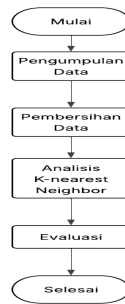
TINJAUAN LITERATUR

Penelitian oleh M. Y. Virasdi dan A. Syaripudin (2023) dengan judul "Penerapan Metode K-Nearest Neighbor untuk Prediksi Penjualan Berbasis Web pada Toko Sembiring" menjelaskan bahwa penggunaan KNN dalam memprediksi penjualan di Toko Sembiring dilakukan untuk meningkatkan efisiensi kerja dari sistem manual. Dengan 120 data dataset yang berisi atribut kuantitas produk, kuantitas terjual, dan harga, diperoleh akurasi sebesar 93,3%, serta sistem pengelolaan data (simpan, edit, hapus, tambah, report) berjalan tanpa error [5]. Penelitian oleh I. G. Gusti, M. Nasrun, dan R. A. Nugrahaeni (2019) dengan judul "Rekomendasi Sistem Pemilihan Mobil Menggunakan K-Nearest Neighbor (KNN) Collaborative Filtering" menunjukkan bahwa metode ini dapat memberikan hasil rekomendasi mobil yang akurat berdasarkan kedekatan data uji dengan data pelatihan. Hasil penelitian menunjukkan bahwa untuk mendapatkan 10 mobil terbaik, jarak maksimal yang digunakan adalah 5%, dan akurasi terbaik diperoleh saat nilai $K = 10$ dengan tingkat akurasi sebesar 95,15% [6]. Penelitian oleh R. Bahtiar (2023) dengan judul "Implementasi Data Mining untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor" membahas penerapan data mining untuk memprediksi penjualan kusen terlaris di Toko Kusen Kembar Djaya. Penelitian bertujuan membantu pemilik toko dalam merencanakan penyediaan stok bahan produksi, dengan hasil akurasi prediksi mencapai 88,89% pada data penjualan dan 80,00% pada data bahan terpakai [7]. Penelitian oleh A. Amalia, A. Zaidiah, dan I. N. Isnainiyah (2023) dengan judul "Prediksi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor" menunjukkan bahwa algoritma KNN dapat digunakan untuk memprediksi kualitas udara berdasarkan Indeks Standar Pencemar Udara (ISPU) di DKI Jakarta. Dengan tujuh parameter kualitas udara (PM_{10} , $PM_{2.5}$, SO_2 , CO , O_3 , NO_2 , dan MAX), KNN mampu mengklasifikasikan data dengan akurasi tinggi. Hasil evaluasi model menunjukkan nilai $K = 7$ memberikan performa terbaik dengan akurasi 96%, presisi 92%, recall 95%, dan f-measure 93%, membuktikan bahwa metode ini efektif untuk membantu pemantauan dan penanggulangan pencemaran udara [8].

Penelitian terakhir oleh G. A. J. Satvika, I. N. Sukajaya, dan I. G. A. Gunadi (2024) dengan judul "Improving K-Nearest Neighbor Performance Using Permutation Feature Importance to Predict Student Success in Study" membahas peningkatan kinerja algoritma KNN dalam memprediksi kelulusan tepat waktu mahasiswa dengan menerapkan teknik Permutation Feature Importance (PFI). Dengan menganalisis data 460 mahasiswa dari program Pendidikan Teknik Informatika Universitas Pendidikan Ganesha periode 2014–2017, hasil penelitian menunjukkan bahwa integrasi PFI meningkatkan akurasi dari 90,22% menjadi 92,39%, presisi dari 50,00% menjadi 62,50%, dan F1-score dari 52,63% menjadi 58,82%. Pendekatan ini menyarankan pemilihan nilai n-shuffle yang optimal pada PFI untuk mendapatkan pola kontribusi fitur yang lebih akurat [9].

METODE PENELITIAN

Metode penelitian adalah bidang ilmu yang mencakup berbagai pendekatan dalam penelitian. Metodologi penelitian membahas konsep-konsep teoritis mengenai berbagai metode. Pada penelitian ini, dibuat sebuah diagram alur yang memaparkan secara singkat mengenai tahapan penelitian yang akan dilakukan. Berikut merupakan diagram alur metodologi penelitian.



Gambar 1. Alur Pemodelan KNN

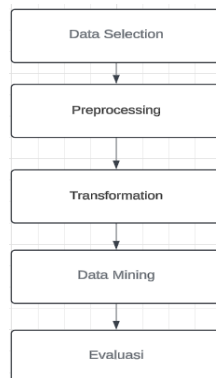
Tahapan Diagram Alur Pemodelan KNN

1. Pengumpulan Data

Pengumpulan data merupakan langkah awal untuk memahami suatu kejadian atau masalah dengan mengekstrak informasi atau data dari beberapa sumber[10]. Metode penelitian dalam studi ini menggunakan pendekatan kuantitatif dengan analisis data sekunder [11]. Data diperoleh dari situs *Gaikindo.com* (tahun, data spesifikasi dan data penjualan) dan *Oto.com* untuk data harga. Penelitian ini menggunakan data dari tahun 2023 hingga 2024.

2. Analisis Data untuk Pemodelan KNN

Alur tahapan dari Analisis Data untuk Pemodelan KNN dapat dilihat dari gambar berikut ini:



Gambar 2. Analisis Data untuk Pemodelan KNN

Pada dataset yang digunakan dalam penelitian ini, terdapat satu variabel terikat (Y) dan variabel bebas (X). Variabel terikat tersebut adalah kategori penjualan, sedangkan variabel bebasnya adalah Kategori Mobil, Brand, Tipe Model, CC, Fuel, Trans, PS/HP, Seater, Tank Capt, Harga dan Jumlah penjualan.

HASIL PENELITIAN

1. Data Yang Digunakan Pada Analisis Data

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai[12]. Data yang digunakan dalam penelitian terjadi antara 2023 hingga 2024 yang berjumlah 234 data:

Tabel 1. Dataset Spesifikasi Mobil

Kategori Mobil	Brand	Tipe Model	CC	Tank Cap	PS/HP	SEATER	Jumlah	Harga	Trans	Fuel
Sedan	HONDA	Honda City	1497	40	120	5	300	402000000	CVT	G
Sedan	BMW	i7 xDrive60 Limousine RHD AT	0	0	544	5	134	3285000000	AT	BEV
Sedan	HYUNDAI	Ioniq 6	0	0	0	5	263	1220000000	AT	BEV

Sedan	MERCEDES BENZ	CLA 200 AMG Line (C118)	1332	43	163	5	265	1200000000	AT	G
Sedan	MERCEDES BENZ	AMG CLA 45 S (C118)	1332	43	163	5	242	1900000000	AT	G
Sedan	MAZDA	Mazda 2 Sedan	1496	44	111	5	507	367000000	AT	G
Sedan	MORRIS GARAGE	5 GT Activate 1.5L	1498	45	114	5	14	347900000	CVT	G
Sedan	MORRIS GARAGE	5 GT Ignite 1.5L	1498	45	115	5	220	377900000	CVT	G
.....										...
4X2 TYPE SALES	CHERY	Omoda E5	1500	61	156	5	4106	498800000	CVT	BEV
4X2 TYPE SALES	CITROEN	All New C3 Aircross SUV	1200	45	110	7	761	289900000	AT	G
4X2 TYPE SALES	DAIHATSU	New Terios IDS X MT (MC 2024)	1496	45	104	7	8582	253500000	MT	G
4X4 TYPE SALES	SUZUKI	Jimny 5 Doors 2Tone AT	1462	40	104	4	1001	253500000	AT	G
PICK UP	DAIHATSU	Gran Max PU 1.3 3WAY	1298	43	88	3	593	163550000	MT	G
PICK UP	TOYOTA	Hilux 2.0 PU SC	2000	65	136	2	606	283700000	MT	G
TRUCK	HINO	136HD	4009	100	130	2	957	465700000	MT	D
TRUCK	MITSUBIS HI FUSO	FE 74 N	3908	100	125	3	841	495000000	MT	D

2. Preprocessing Data

Data awal berjumlah 234 dibagi menjadi 213 data train dan 21 data test untuk diberi encoding dan diberi label kategori.

- Membuat label encoding dan mapping untuk variabel trans, fuel, kategori mobil dan kategori penjualan. Fungsi "encoding" digunakan untuk mengubah kata string menjadi angka (numerik) dan fungsi mapping untuk menampilkan kata yang telah menjadi numerik. Contoh penggunaannya sebagai berikut.

```
# Inisialisasi LabelEncoder
label_encoder = LabelEncoder()

# Fit label encoder pada trans untuk memastikan mapping konsisten
combined_trans = pd.concat([A1['trans'], A2['trans']])
label_encoder.fit(combined_trans)

A1['trans'] = label_encoder.transform(A1['trans'])
A2['trans'] = label_encoder.transform(A2['trans'])

# Simpan mapping untuk trans
mapping_trans = dict(zip(label_encoder.classes_, label_encoder.transform(label_encoder.classes_)))
print("Mapping for 'trans':", mapping_trans)

# Fit label encoder pada fuel
combined_fuel = pd.concat([A1['fuel'], A2['fuel']])
label_encoder.fit(combined_fuel)

A1['fuel'] = label_encoder.transform(A1['fuel'])
A2['fuel'] = label_encoder.transform(A2['fuel'])

# Simpan mapping untuk fuel
mapping_fuel = dict(zip(label_encoder.classes_, label_encoder.transform(label_encoder.classes_)))
print("Mapping for 'fuel':", mapping_fuel)
```

Mapping for 'trans': {'AT': 0, 'CVT': 1, 'MT': 2}
Mapping for 'fuel': {'BEV': 0, 'D': 1, 'EV': 2, 'G': 3, 'HEV': 4, 'HYBRID': 5}

Mapping for 'Kategori Mobil': {'4X2 TYPE SALES': 0, '4X4 TYPE SALES': 1, 'DOUBLE CABIN 4X2 / 4X4': 2, 'PICK UP': 3, 'Sedan': 4, 'TRUCK': 5}
Mapping for 'Kategori Penjualan': {'cukup laris': 0, 'kurang laris': 1, 'sangat laris': 2}

Gambar 3. Label Encoder dan Mapping

Gambar 3. menunjukkan hasil proses encoding menggunakan Label Encoder, data kategorikal dikonversi menjadi nilai numerik untuk memudahkan pemrosesan dalam analisis dan pemodelan machine learning. Pada variabel transmisi, tipe "AT" dikonversi menjadi angka 0, tipe "CVT" menjadi angka 1, dan tipe "MT" menjadi angka 2. Kemudian, pada variabel bahan bakar, jenis bahan bakar "BEV" diberikan angka 0, jenis bahan bakar "D" diberikan angka 1, jenis bahan bakar "EV" diberikan angka 2, jenis bahan bakar "G" diberikan angka 3, jenis bahan bakar "HEV" diberikan angka 4, dan jenis bahan bakar "HYBRID" diberikan angka 5. Selanjutnya, pada variabel kategori mobil, tipe mobil "4X2 TYPE SALES" dikodekan menjadi angka 0, tipe mobil "4X4 TYPE SALES" dikodekan menjadi angka 1, tipe mobil "DOUBLE CABIN 4X2 / 4X4" dikodekan menjadi angka 2, tipe mobil "PICK UP" dikodekan menjadi angka 3, tipe mobil "Sedan" dikodekan menjadi angka 4, dan tipe mobil "TRUCK" dikodekan menjadi angka 5. Sementara itu, pada variabel kategori penjualan, kategori penjualan "cukup laris" dikodekan menjadi angka 0, kategori penjualan "kurang laris" dikodekan menjadi angka 1, dan kategori penjualan "sangat laris" dikodekan menjadi angka 2.

b. Memberi Label Kategori

Memberikan label kategori Sangat Laris, Cukup laris dan Kurang Laris sebelum masuk ke proses transformasi.

```
[ ] # 1. Buat kategori 'sangat laris', 'cukup', dan 'kurang laris'
def categorize_sales(x):
    if x > 450:
        return 'sangat laris'
    elif x > 76:
        return 'cukup laris'
    else:
        return 'kurang laris'
di['kategori_penjualan'] = di['Jumlah'].apply(categorize_sales)
```

Gambar 4. Label Kategori Penjualan

Penjelasan Gambar 4. menunjukkan kategori penjualan berdasarkan jumlah unit yang terjual dalam satu tahun. Produk dengan penjualan ≤ 76 unit dikategorikan sebagai "Kurang Laris," $76 < Y \leq 450$ sebagai "Cukup Laris," dan > 450 unit sebagai "Sangat Laris." Kategori ini ditentukan dengan menghitung rentang dari nilai minimal hingga maksimal penjualan.

c. Transformation

Transformation adalah tahapan data diubah dalam bentuk sesuai kebutuhan [13]. Pada tahap ini dilakukan pembagian data train dan data test yang telah di encoding dan diberikan label kategori. Pembagian dilakukan menjadi 90/10 yaitu 90% untuk data train dan 10% untuk data test atau 213 data train dan 21 data test.

```
[ ] yh = pd.read_excel('modified_sales_data.xlsx') #yh = data train
yt = pd.read_excel('modified_sales_data_test.xlsx') #yt = data test
```

Gambar 5. Tahap Transformation

d. Data Mining

Data Mining adalah serangkaian proses yang dilakukan agar dapat menggali informasi atau pengetahuan yang awalnya tidak diketahui dari data yang besar, pengetahuan yang didapatkan dapat diperoleh dengan melakukan ekstraksi serta memilih atribut penting dari data dan berpengaruh terhadap prediksi dari data yang besar yang diproses dengan menggunakan teknik statistik, kecerdasan buatan dan machine learning [14]. Pada tahap ini memiliki beberapa tahapan seperti menampilkan hasil jarak euclidean, menentukan fitur independen dan target, normalisasi data, cari nilai k optimal, latih dan evaluasi model knn, menampilkan hasil pie chart perbandingan data aktual dan prediksi, uji model dengan data baru dan terakhir tahap analisis feature importance.

```

kategori_penjualan  distance_to_s1_first_row
3                   0      6.480000e+08
1                   0      3.647864e+09
2                   0      8.354783e+08
3                   0      8.119754e+08
4                   0      1.722354e+09
..                  ...      ...
16                  2      8.102250e+08
17                  2      9.179608e+08
18                  2      7.753777e+08
19                  2      5.877620e+08
20                  2      5.627379e+08
[234 rows x 13 columns]
```

Gambar 6. hasil jarak euclidean

```
#menentukan fitur (variabel independen) dan target (variabel dependen)
features = ['Kategori Mobil', 'CC', 'Tank Cap', 'PS/HP', 'SEATER', 'trans', 'fuel', 'Harga']
X_train = yh[features]
Y_train = yh['kategori_penjualan']
```

Gambar 7. sintaks fitur independen dan target

Penjelasan :

Fitur code (X_train dan X_test) adalah pemisahan fitur independen yang akan digunakan untuk prediksi. Dalam fitur Code (Y_train dan Y_test) untuk menentukan target label kategori penjualan yang akan diprediksi.

```
#Normalisasi fitur (KNN sensitif terhadap skala fitur)
scaler_X = StandardScaler()
X_train_scaled = scaler_X.fit_transform(X_train)
X_test_scaled = scaler_X.transform(X_test)
```

Gambar 8. sintaks Normalisasi Data (Standarisasi)

```
numNeighbors = range(1, 20)
trainAcc = []
testAcc = []

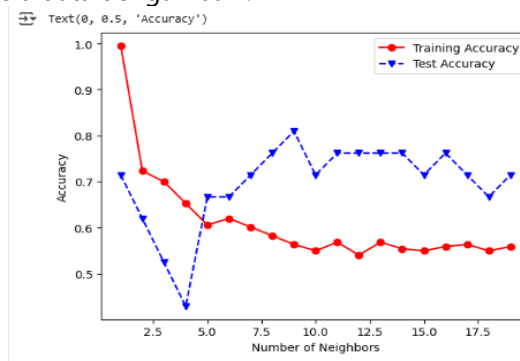
for k in numNeighbors:
    clf = KNeighborsClassifier(n_neighbors=k, metric='minkowski', p=2)
    clf.fit(X_train, Y_train)
    Y_predTrain = clf.predict(X_train)
    Y_predTest = clf.predict(X_test)
    trainAcc.append(accuracy_score(Y_train, Y_predTrain))
    testAcc.append(accuracy_score(Y_test, Y_predTest))

plt.plot(numNeighbors, trainAcc, 'ro-', numNeighbors, testAcc, 'bv--')
plt.legend(['Training Accuracy', 'Test Accuracy'])
plt.xlabel('Number of Neighbors')
plt.ylabel('Accuracy')
```

Gambar 9. Sintaks Pelatihan Model KNN

Penjelasan :

Optimasi nilai K pada K-NN menjadi kunci dalam meningkatkan performa klasifikasi. Nilai K yang terlalu kecil rentan terhadap outlier, sedangkan nilai K yang terlalu besar dapat menyebabkan kesalahan klasifikasi akibat perhitungan jarak yang tidak optimal[15]. Pada tahap ini, kita mencoba berbagai jumlah tetangga terdekat (K) agar dapat melihat bagaimana pengaruhnya terhadap akurasi model yang dibuat. Untuk mengukur performa model, kita menyiapkan dua daftar kosong, yaitu train Acc menyimpan akurasi data training dan test Acc untuk menyimpan akurasi data testing. Setelah model selesai dilatih, kita menggunakannya untuk memprediksi data training dan testing. Prediksi pada data training dilakukan untuk melihat apakah model dapat dikenali, sedangkan prediksi pada data testing digunakan untuk menguji seberapa baik model yang dapat mengenali data baru yang belum pernah dilihat atau dicoba sebelumnya. Melalui proses ini, kita dapat menemukan jumlah tetangga terbaik agar model KNN bekerja secara maksimal dalam mengenali pola data dengan baik.



Gambar 10. Tetangga Optimal K

Penjelasan :

Gambar menunjukkan bahwa grafik data train dan data test berdekatan yang menandakan

model sudah cukup baik dan kedua grafik berdekatan pada nilai $k=5$.

```
[10] # Latih model KNN
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(X_train_scaled, Y_train)
```

KNeighborsClassifier

KNeighborsClassifier()

Gambar 11. Sintaks Model K-NN

Penjelasan :

Fitur `knn = KNeighborsClassifier(n_neighbors=5)` baris tersebut membuat objek model KNN dengan jumlah tetangga terdekat (k) adalah 5. Parameter ini menentukan jumlah tetangga terdekat yang digunakan dalam proses klasifikasi. Data ini kemudian dibagi menjadi data latih dan data uji menggunakan fungsi `train_test_split` untuk memudahkan pelatihan dan evaluasi model [16].

```
# Predict on the test set
Y_pred = knn.predict(X_test_scaled)

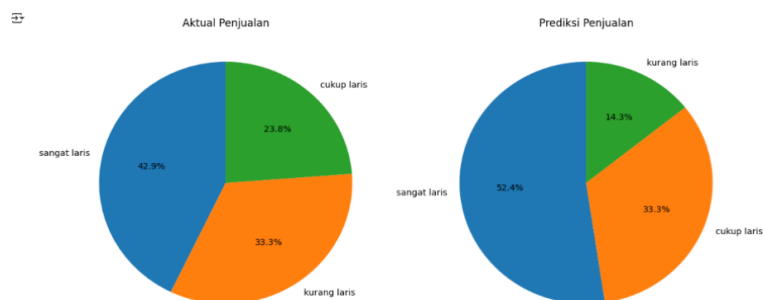
# Evaluate the model
accuracy = accuracy_score(Y_test, Y_pred)
print(f"Accuracy of the KNN model: {accuracy}")
```

Accuracy of the KNN model: 0.7142857142857143

Gambar 12. Sintaks Accuracy Score model K-NN

Penjelasan :

Hasil akurasi model K-NN sebesar 0,71 dengan nilai $k = 5$, yang menunjukkan bahwa model memiliki performa yang cukup baik dalam melakukan prediksi.



Gambar 13. Visualisasi Plot Garis untuk Perbandingan Prediksi

Penjelasan:

Hasil visualisasi menunjukkan perbandingan antara data aktual dan prediksi penjualan berdasarkan tiga kategori: sangat laris, cukup laris, dan kurang laris. Dalam data aktual, kategori "sangat laris" memiliki porsi terbesar, yaitu 42.9%, diikuti oleh "kurang laris" sebesar 33.3%, dan "cukup laris" sebesar 23.8%. Sementara itu, dalam hasil prediksi, kategori "sangat laris" meningkat menjadi 52.4%, "cukup laris" tetap pada 33.3%, sedangkan "kurang laris" menurun menjadi 14.3%.

```
#prediksi jumlah
databaru = {'Kategori Mobil': 0, 'CC': 1329, 'Tank Capt': 43, 'PS/HP': 97, 'SEATER': 7, 'Harga': 245550000, 'trans': 0, 'fuel': 3}
databaru = pd.DataFrame([databaru])
X_new = databaru[features]
X_new_scaled = scaler_X.transform(X_new)
y_pred_new = knn.predict(X_new_scaled)

print(f'prediksi jumlah: {y_pred_new[0]}')
```

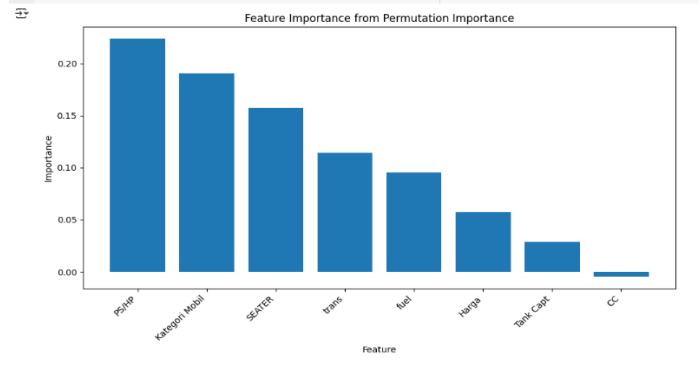
prediksi jumlah: 2

Gambar 14. Sintaks Pada Data Uji Baru

Penjelasan:

Sintaks berfungsi mencoba memasukkan data baru dan melihat hasil prediksi kategori klasifikasi. Hasil menunjukkan prediksi jumlah 2 memiliki arti bahwa mobil dengan

spesifikasi tersebut termasuk dalam kategori Sangat Laris.

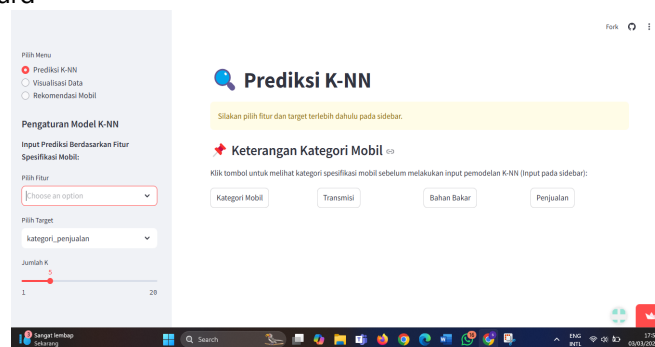


Gambar 15. Visualisasi untuk Pengaruh Faktor

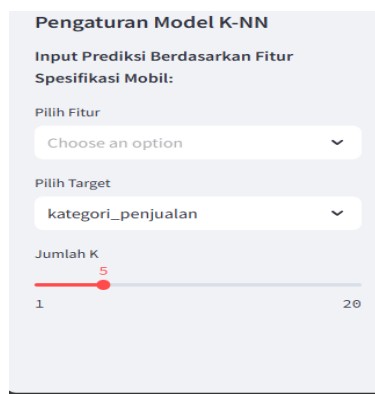
Penjelasan:

Visualisasi menunjukkan bahwa dalam pemilihan mobil, konsumen pertama-tama mempertimbangkan variabel PS/HP, diikuti oleh kategori mobil, jumlah seater, transmisi, jenis bahan bakar (fuel), harga, kapasitas tangki (tank capt), dan terakhir CC.

e. Dashboard



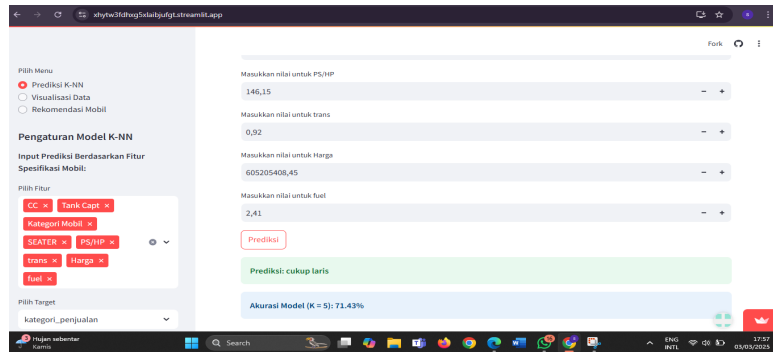
Gambar 16. Tampilan Awal Dashboard



Gambar 17. Input Pemodelan K-nearest Neighbour

Penjelasan:

Fitur pemodelan K-NN terletak di sidebar dashboard sebagai input pemodelan K-NN, kolom fitur berperan sebagai fitur utama, dan kolom target berfungsi sebagai target. Kolom fitur berisi spesifikasi mobil, seperti kapasitas mesin (CC), jenis bahan bakar, sistem transmisi, kapasitas tangki bensin, kategori mobil, dan harga, sementara kolom target berisi kategori penjualan. Fitur nilai K berfungsi sebagai parameter filter memungkinkan pengguna memilih nilai K dalam rentang 1 hingga 20. Nilai k digunakan mencari jumlah mayoritas dari target atau label data training, sehingga nilai k tidak boleh lebih dari jumlah record dataset training[17].



Gambar 18. Hasil Pemodelan K-nearest Neighbour

Penjelasan:

Prinsip kerja K-Nearest Neighbor adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (Neighbor) terdekatnya dalam data pelatihan[18]. Setelah pengguna memasukkan nilai yang sesuai untuk diprediksi, mereka dapat menekan tombol "Prediksi" untuk menampilkan hasil prediksi.

PEMBAHASAN

Penelitian menganalisis faktor-faktor yang mempengaruhi penjualan mobil berdasarkan model klasifikasi menggunakan metode k- nearest neighbor. Hasil analisis menunjukkan bahwa model knn mendapatkan accuracy score sebesar 0,71 yang menunjukkan bahwa model knn untuk prediksi kategori klasifikasi memiliki tingkat akurasi yang cukup baik. Hasil visualisasi menunjukkan perbandingan antara data aktual dan prediksi yaitu dalam data aktual, kategori "sangat laris" memiliki porsi terbesar, yaitu 42.9%, diikuti oleh "kurang laris" sebesar 33.3%, dan "cukup laris" sebesar 23.8%. Lalu untuk hasil prediksi, kategori "sangat laris" meningkat menjadi 52.4%, "cukup laris" tetap pada 33.3%, sedangkan "kurang laris" menurun menjadi 14.3%. Hasil percobaan menggunakan data baru menunjukkan bahwa model klasifikasi dapat digunakan dengan cukup baik untuk memprediksi kategori klasifikasi pada penjualan mobil. Hasil dari fitur *Feature Importance* menunjukkan faktor-faktor yang mempengaruhi konsumen dalam pembelian mobil menunjukkan bahwa konsumen pertama-tama mempertimbangkan variabel PS/HP, diikuti oleh kategori mobil, jumlah seater, transmisi, jenis bahan bakar (fuel), harga, kapasitas tangki (tank capt), dan terakhir CC. Hasil analisis tersebut bisa menjadi acuan bagi produsen dalam merancang dan memproduksi mobil yang sesuai dengan tren dan preferensi pasar konsumen.

KESIMPULAN

Hasil analisis menunjukkan bahwa proses pemodelan klasifikasi K-NN yang telah dianalisis, diperoleh akurasi model K-NN sebesar 0,71 dengan nilai $k = 5$, yang menunjukkan bahwa model memiliki performa yang cukup baik dalam melakukan prediksi penentuan klasifikasi kategori penjualan mobil. Faktor mempengaruhi penjualan mobil dianalisis menggunakan data uji dengan pemodelan K-NN, dimana data latih dan data uji digunakan untuk mengidentifikasi faktor utama yang mempengaruhi keputusan konsumen pada penjualan mobil. Faktor yang paling berpengaruh adalah faktor variabel PS/HP, diikuti oleh kategori mobil, jumlah seater, transmisi, jenis bahan bakar (fuel), harga, kapasitas tangki (tank capt), dan terakhir CC.

Hasil penerapan adalah melihat kategori klasifikasi kategori penjualan mobil berdasarkan faktor-faktor berpengaruh menggunakan model klasifikasi KNN. Selain itu, hasil membantu produsen dalam merancang dan memproduksi mobil yang sesuai dengan tren dan preferensi pasar konsumen masa mendatang.

REFERENCES

- [1] D. T. R. R. A. S. Muhammad Nabil Afkar, "PREDIKSI PENJUALAN MOBIL TOYOTA DI INDONESIA MENGGUNAKAN MULTI-LAYERPERCEPTRON," *JIP (Jurnal Informatika Polinema)*, vol. XI, no. 1, pp. 91-98, 2024.

- [2] R. K. I. S. H. W. F. Amalia, "Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Kemampuan Lulusan Siswa Dalam Bersaing untuk Mendapatkan Pekerjaan," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. I, pp. 1-8, 2024.
- [3] A. F. A. A. J. A. S. A. O. E. P. E. R. Z. A. S. F. Mayang Fadillah Haryanti, "Pengaruh Data Mining, Strategi Perusahaan Terhadap Laporan Kinerja Perusahaan," *Jurnal Manajemen dan Bisnis*, vol. III, pp. 71-90, 2024.
- [4] M. Fansyuri, "ANALISA ALGORITMA KLASIFIKASI K-NEAREST NEIGHBOR DALAM MENENTUKAN NILAI AKURASI TERHADAP KEPUASAN PELANGGAN," *Humanika: Jurnal Ilmu Sosial, Pendidikan, dan Humaniora*, vol. I, pp. 29-33, 2020.
- [5] A. S. Muhammad Yusuf Virasdi, "Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Berbasis Web Pada Toko Sembiring," *OKTAL : Jurnal Ilmu Komputer dan Science*, vol. II, pp. 2822-2832, 2023.
- [6] M. N. R. A. N. Ilham Gumantung Gusti, "REKOMENDASI SISTEM PEMILIHAN MOBIL MENGGUNAKAN K-NEAREST NEIGHBOR (KNN)," *Jurnal TEKTRIKA*, vol. IV, pp. 26-32, 2019.
- [7] R. Bahtiar, "Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor (Studi Kasus : Toko Kusen Kembar Djaya)," *Jurnal Informatika Universitas Pamulan*, vol. I, no. 3, pp. 1-12, 2023.
- [8] A. Z. I. N. I. Adinda Amalia, "Prediksi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. VII, no. 2, p. 496-507, 2022.
- [9] I. N. S. I. G. A. G. Gd. Aditya Jana Satvika, "G. A. J. Satvika, I. N. Sukajaya, and I. G. A. Gunadi, "Improving K-Nearest Neighbor Performance Using Permutation Feature Importance to Predict Student Success in Study," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. XXXV, no. 3, pp. 1835-1844, 2024.
- [10] N. R. Y. M. R. H. Kiki Mustaqim, "Perbandingan Metode Regresi Linear Dan K-Nearest Neighbor (KNN) Dalam Memprediksi Produksi Tanaman Padi Di Pulau Sumatera," *DSI: Jurnal Data Science Indonesia*, vol. IV, no. 2, pp. 60-74, 2025.
- [11] F. A. M. S. S. A. P. P. N. Noviana Riza, "IMPLEMENTASI REGRESI LINEAR BERGANDA PREDIKSI FAKTOR-FAKTOR INDEKS PEMBANGUNAN MANUSIA DI PROVINSI JAWABARAT," *Masaliq Journal*, vol. V, no. 1, pp. 69-89, 2024.
- [12] A. N. Syaiful Zuhri Harahap, "TEKNIK DATA MINING UNTUK PENENTUAN PAKET HEMAT SEMBAKO DAN KEBUTUHAN HARIAN DENGAN MENGGUNAKAN ALGORITMA FP-GROWTH (STUDI KASUS DI ULFAMART LUBUK ALUNG)," *Informatika : Jurnal Ilmiah Fakultas Sains dan Teknologi, Universitas Labuhanbatu*, vol. VII, no. 3, pp. 111-119, 2019.
- [13] D. J. Fauzan Alghifari, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *JURNAL ILMIAH INFORMATIKA*, vol. IX, no. 2, pp. 76-81, 2021.
- [14] A. H. Z. F. Ahmad Yogianto, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," *G-Tech: Jurnal Teknologi Terapan*, vol. VIII, no. 3, pp. 1720-1728, 2024.
- [15] G. H. F. A. P. F. A. D. S. I. G. R. Nur Tulus Ujianto, "Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis," *IT-EXPLORE Jurnal Penerapan Teknologi Informasi dan Komunikasi*, vol. II, no. 2, pp. 33-43, 2023.
- [16] A. P. N. F. C. A. N. B. H. A. S. Muhammad Rizky Maulana, "KNN-Based Handwritten Digit Classification with Accuracy Analysis and Visualization," *JURNAL ILMIAH SISTEM INFORMASI (JUSI)*, vol. IV, no. 2, pp. 142-150, 2025.
- [17] M. A. F. B. aiful Ulya, "S. Ulya, M. A. Soeleman, and F. Budiman, "Optimasi Parameter K Pada Algoritma K-NN Untuk Klasifikasi Prioritas Bantuan Pembangunan Desa," *Techno.COM*, vol. XX, no. 1, pp. 83-96, 2021.
- [18] Z. R. M. F. Zikri Hadiansyah, "Implementasi Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Kanker Paru Paru," *Journal of Computer System and Informatics (JoSYC)*, vol. VI, no. 1, pp. 96-106, 2024.