Universitas Potensi Utama

cek plagiasi KTI Dimas Sembiring

Editor Publidimas
publidimas
Universitas 17 Agustus 1945 Semarang

Document Details

Submission ID
trn:oid:::1:3479451121

Submission Date
Feb 12, 2026, 9:56 AM GMT+7

Download Date
Feb 12, 2026, 9:57 AM GMT+7

File Name
Journal_DIMAS_SEMBIRING_-_Dimas_Sembiring.docx

File Size
5.8 MB

10 Pages

2,390 Words

15,975 Characters

30% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Top Sources

29% Internet sources
20% Publications
16% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

29% Internet sources
20% Publications
16% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	
	jurnal.dharmawangsa.ac.id	8%
2	Internet	
	www.ijcaonline.org	3%
3	Internet	
	digilib.uin-suka.ac.id	2%
4	Internet	
	www.globalscientificjournal.com	2%
5	Student papers	
	University of Cincinnati	1%
6	Internet	
	kti.potensi-utama.org	1%
7	Internet	
	arxiv.org	<1%
8	Internet	
	as-proceeding.com	<1%
9	Internet	
	editorapantanal.com.br	<1%
10	Internet	
	journal.aptii.or.id	<1%
11	Student papers	
	Asia Pacific International College	<1%

12	Internet	
jutif.if.unsoed.ac.id		<1%
13	Student papers	
Universitas 17 Agustus 1945 Semarang		<1%
14	Internet	
journal.dharmawangsa.ac.id		<1%
15	Student papers	
La Trobe University		<1%
16	Internet	
lseee.net		<1%
17	Publication	
You-Jeng Chang, Ying-Lei Lin, Ping-Feng Pai. "Support Vector Machines with Hype...		<1%
18	Internet	
researchhub.id		<1%
19	Internet	
repository.upi.edu		<1%
20	Publication	
Fang Fang, Jialiang Li, Xiaochao Xia. "Semiparametric model averaging prediction...		<1%
21	Publication	
Indah Clara Sari. "INTEGRASI MODEL DEEP LEARNING EFFICIENTNET-B0 UNTUK D...		<1%
22	Internet	
e-journal.potensi-utama.ac.id		<1%
23	Internet	
zombiedoc.com		<1%
24	Internet	
media.neliti.com		<1%
25	Internet	
www.coursehero.com		<1%

26 Internet

journal.maranatha.edu <1%

27 Internet

mardiya.wordpress.com <1%

28 Internet

text-id.123dok.com <1%

29 Internet

www.marketeers.com <1%

30 Internet

www.semanticscholar.org <1%

31 Publication

M. Dimas Ubaidilah, Ahmad Tohari, Sherla Dian Mutia, Muhammad Ferri Harmaw... <1%

Djtechno : Jurnal Teknologi Informasi
Vol. xxx, No. xxx, Bulan Tahun
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

IMPLEMENTASI ALGORITMA *LOGISTIC REGRESSION* DALAM MEMPREDIKSI KELAS HARGA *SMARTPHONE* BERDASARKAN SPESIFIKASI

Dimas Sembiring[1], Mas Ayoe Nasution [2]
1) Program Studi Informatika UPU, Jl. KL. Yos Sudarso Km 6,5 No. 3A, Medan, 20241, Indonesia
2) Dosen Program Studi Informatika UPU, Jl. KL. Yos Sudarso Km 6,5 No. 3A, Medan, 20241, Indonesia

Article Info

Article history:

Received: dd, month, yyyy
Revised: dd, month, yyyy
Accepted: dd, month, yyyy

ABSTRACT (10PT)

Abstrak

Perkembangan industri smartphone yang semakin cepat menuntut pemahaman lebih mendalam mengenai segmentasi harga berdasarkan spesifikasi perangkat. Penelitian ini bertujuan membangun model klasifikasi harga smartphone ke dalam beberapa kategori menggunakan algoritma *Logistic Regression*, serta mengevaluasi pengaruh *preprocessing* dan *hyperparameter tuning* terhadap kinerja model. Hasil menunjukkan bahwa model mampu mengklasifikasikan harga dengan akurasi yang baik, dengan RAM dan resolusi layar sebagai fitur paling berpengaruh. Normalisasi data juga terbukti meningkatkan kestabilan prediksi.

Kata Kunci: ***Regresi Logistik, Klasifikasi Harga, Smartphone, Normalisasi Data, Penyetelan Hyperparameter***

Abstract

The rapid growth of the smartphone industry requires a deeper understanding of price segmentation based on device specifications. This study aims to develop a smartphone price classification model using Logistic Regression and to evaluate the impact of data preprocessing and hyperparameter tuning on model performance. The results show that the model achieves good classification accuracy, with RAM and screen resolution as the most influential features. Data normalization also improves the stability of the model's predictions.

Keywords: Logistic Regression, Price Classification, Smartphone, Data Normalization, Hyperparameter Tuning

Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercialL ShareAlike 4.0 (CC-BY-NC-SA).

Corresponding Author:
Pilih penulis yang akan menjadi korespondensi author
E-mail : corresponding author, Nomor whatsapp (akan di hapus jika masuk proses revieuw)

Djtechno : Jurnal Teknologi Informasi
Vol. xxx, No. xxx, Bulan Tahun
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

1. PENDAHULUAN

Dalam era digital yang semakin berkembang pesat, smartphone telah menjadi kebutuhan utama bagi masyarakat di berbagai kalangan. Perangkat ini tidak hanya digunakan untuk komunikasi, tetapi juga untuk pekerjaan, hiburan, pendidikan, dan kegiatan sehari-hari lainnya. Permintaan terhadap smartphone terus meningkat, sehingga mendorong produsen untuk bersaing dalam merilis berbagai varian produk dengan spesifikasi teknis yang beragam, seperti kapasitas RAM, kualitas kamera, ukuran layar, prosesor, daya baterai, dan dukungan jaringan. Namun, keberagaman spesifikasi tersebut juga diiringi dengan variasi harga yang cukup signifikan, mulai dari kelas entry-level, menengah, hingga flagship. Hal ini menimbulkan kebingungan bagi konsumen yang ingin menentukan pilihan berdasarkan kebutuhan dan anggaran yang tersedia. [1]

Bagi pelaku industri, pemahaman mengenai hubungan antara spesifikasi teknis dan kelas harga menjadi hal penting untuk mendukung segmentasi pasar, penetapan harga yang kompetitif, serta pengembangan produk yang tepat sasaran. Oleh karena itu, diperlukan pendekatan analitik yang mampu mengklasifikasikan harga smartphone berdasarkan fitur teknis secara akurat dan efisien.[2]

Machine learning dapat dimanfaatkan untuk menjawab kebutuhan tersebut, salah satunya melalui algoritma Logistic Regression. Metode ini memodelkan probabilitas suatu data masuk ke dalam kategori tertentu dan memiliki keunggulan dari sisi kecepatan, kemudahan interpretasi, serta performa yang baik jika didukung preprocessing seperti normalisasi.[3]

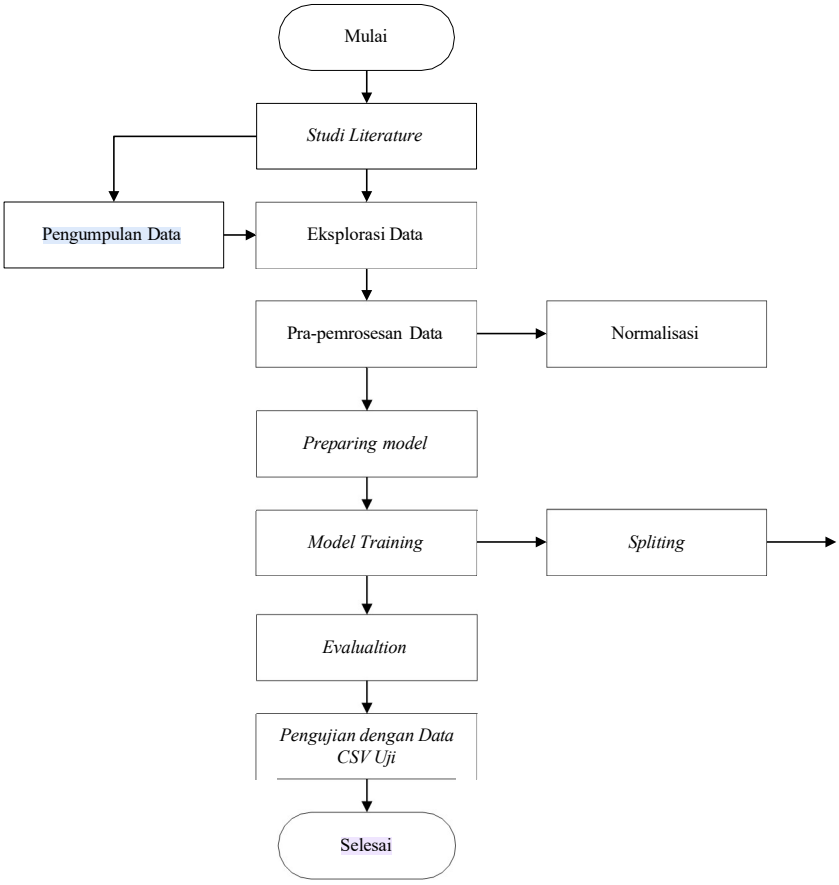
Beberapa penelitian terbaru juga menunjukkan efektivitas Logistic Regression dalam klasifikasi harga smartphone. Ardelia et al. (2024) melaporkan akurasi hingga 98% melalui tuning Grid SearchCV, sementara MeetMariam (2025) memperoleh akurasi 95% setelah feature selection dan hyperparameter tuning. Selain itu, Subhiksha dkk. (2020) menyatakan bahwa kombinasi Logistic Regression dengan metode lain dapat meningkatkan akurasi prediksi hingga lebih dari 80%.[4]

Berdasarkan hal tersebut, penelitian ini berfokus pada penerapan Logistic Regression untuk membangun model klasifikasi harga smartphone berdasarkan spesifikasi teknis. Penelitian ini diharapkan memberikan kontribusi dalam pengembangan kajian machine learning serta membantu konsumen dan industri memahami segmentasi harga secara lebih objektif, sekaligus membuka peluang penerapan dalam sistem rekomendasi dan pendukung keputusan di platform e-commerce.[5]

Djtechno : Jurnal Teknologi Informasi
Vol. xxx, No. xxx, Bulan Tahun
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

2. METODE PENELITIAN

Tahapan yang dilakukan dalam penelitian ini menggunakan pendekatan *kuantitatif eksperimental* yang digambarkan dengan menggunakan tabel Dataset. Beberapa tahapan yang digunakan dalam Dataset adalah sebagai berikut:



Gambar 1. Rancangan Penelitian

3. HASIL DAN PEMBAHASAN

a. Pembahasan

Proses analisis dilakukan berdasarkan data spesifikasi teknis smartphone yang diperoleh dari dataset *Mobile Price Classification*. Hasil yang ditampilkan mencakup pelatihan model menggunakan algoritma *Logistic Regression*, evaluasi performa model, serta pengujian terhadap data uji yang belum diberi label (unlabeled data).

1. *Importing Libs*

Tahapan awal dalam implementasi model klasifikasi harga smartphone berbasis *Logistic Regression* dilakukan dengan mengimpor pustaka (library) yang diperlukan dalam proses analisis dan pengolahan data menggunakan bahasa pemrograman Python. Pustaka-pustaka ini diimpor untuk mendukung berbagai fungsi penting dalam penelitian, mulai dari manipulasi data, visualisasi, normalisasi, pemisahan data latih dan uji, hingga pembangunan dan pelatihan model klasifikasi.

turnitin Page 9 of 15 - Integrity Submission Submission ID trn:oid:::1:3479451121

Djtechno : Jurnal Teknologi Informasi
Vol. xxx, No. xxx, Bulan Tahun
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

```
[1] # Data Analysis
    import pandas as pd
    import matplotlib.pyplot as plt
    import seaborn as sns
    import plotly.express as px
    #Preprocessing---------------------------------------------
    from sklearn.preprocessing import MinMaxScaler,RobustScaler,OneHotEncoder, LabelEncoder
    # Train test Preparation----------------------------
    from sklearn.model_selection import train_test_split
    # Model--------------------------
    from sklearn.linear_model import LogisticRegression
```

Gambar 2. *Import Libararies*

Pustaka pandas digunakan sebagai alat utama untuk membaca dan memanipulasi data dalam bentuk tabel (dataframe), sedangkan matplotlib, seaborn, dan plotly.express digunakan untuk visualisasi data dalam bentuk grafik, diagram, maupun heatmap korelasi. Untuk keperluan pra-pemrosesan data, seperti normalisasi dan pengkodean fitur, digunakan pustaka sklearn.preprocessing dengan fungsi-fungsi seperti MinMaxScaler, RobustScaler, OneHotEncoder, dan LabelEncoder.

Selanjutnya, untuk membagi data menjadi training set dan testing set, digunakan train_test_split dari modul sklearn.model_selection. Terakhir, model klasifikasi Logistic Regression yang menjadi inti dari penelitian ini diakses melalui LogisticRegression dari modul sklearn.linear_model. Seluruh pustaka ini perlu diimpor terlebih dahulu sebelum proses pemodelan dimulai, agar setiap langkah analisis data dan pelatihan model dapat berjalan secara optimal dan efisien dalam lingkungan pemrograman Python yang digunakan.

2. *Data Exploration*

Tahap eksplorasi data bertujuan untuk memahami karakteristik dasar dari dataset yang digunakan, menemukan pola awal, serta mengidentifikasi potensi outlier atau ketidakwajaran data. Berikut adalah penjelasan masing-masing proses eksplorasi data yang dilakukan:

1. Menampilkan Dataset Awal

Perintah df digunakan untuk menampilkan isi awal dataset guna memastikan data telah berhasil dimuat dan memeriksa struktur baris dan kolom secara umum.

```
[3] df
```

	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	n_cores	...	px_height	px_width	ram	sc_h	sc_w	talk_time	three_g	touch_screen	wifi	price_range
0	842	0	2.2	0	1	0	7	0.6	188	2	...	20	756	2549	9	7	19	0	0	1	1
1	1021	1	0.5	1	0	1	53	0.7	136	3	...	905	1988	2631	17	3	7	1	1	0	2
2	563	1	0.5	1	2	1	41	0.9	145	5	...	1263	1716	2603	11	2	9	1	1	0	2
3	615	1	2.5	0	0	0	10	0.8	131	6	...	1216	1786	2769	16	8	11	1	0	0	2
4	1821	1	1.2	0	13	1	44	0.6	141	2	...	1208	1212	1411	8	2	15	1	1	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1995	794	1	0.5	1	0	1	2	0.8	106	6	...	1222	1890	668	13	4	19	1	1	0	0
1996	1965	1	2.6	1	0	0	39	0.2	187	4	...	915	1965	2032	11	10	16	1	1	1	2
1997	1911	0	0.9	1	1	1	36	0.7	108	8	...	868	1632	3057	9	1	5	1	1	0	3
1998	1512	0	0.9	0	4	1	46	0.1	145	5	...	336	670	869	18	10	19	1	1	1	0
1999	510	1	2.0	1	5	1	45	0.9	168	6	...	483	754	3919	19	4	2	1	1	1	3

2000 rows × 21 columns

Gambar 3. Dataset Awal

turnitin Page 9 of 15 - Integrity Submission Submission ID trn:oid:::1:3479451121

Djtechno : Jurnal Teknologi Informasi
Vol. xxx, No. xxx, Bulan Tahun
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

2. Pemeriksaan Struktur Data

Fungsi df.info() digunakan untuk menampilkan informasi struktur dataset, termasuk jumlah data per kolom, tipe data (integer/float), serta pengecekan apakah ada nilai yang hilang (missing values).

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 21 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   battery_power  2000 non-null   int64
 1   blue           2000 non-null   int64
 2   clock_speed    2000 non-null   float64
 3   dual_sim       2000 non-null   int64
 4   fc             2000 non-null   int64
 5   four_g         2000 non-null   int64
 6   int_memory     2000 non-null   int64
 7   m_dep          2000 non-null   float64
 8   mobile_wt      2000 non-null   int64
 9   n_cores        2000 non-null   int64
 10  pc             2000 non-null   int64
 11  px_height      2000 non-null   int64
 12  px_width       2000 non-null   int64
 13  ram            2000 non-null   int64
 14  sc_h           2000 non-null   int64
 15  sc_w           2000 non-null   int64
 16  talk_time      2000 non-null   int64
 17  three_g        2000 non-null   int64
 18  touch_screen   2000 non-null   int64
 19  wifi           2000 non-null   int64
 20  price_range    2000 non-null   int64
dtypes: float64(2), int64(19)
memory usage: 328.3 KB
```

Gambar 4. Informasi Struktur Dataset

Berdasarkan hasil keluaran dari fungsi df.info(), dapat diketahui bahwa dataset yang digunakan dalam penelitian ini terdiri dari 2.000 entri (baris data) dan 21 kolom fitur. Setiap baris merepresentasikan satu unit data smartphone yang memiliki spesifikasi teknis tertentu serta kategori harga yang sesuai. Keseluruhan kolom dalam dataset ini tidak mengandung nilai kosong (null), yang ditunjukkan dengan jumlah *non-null count* sebanyak 2000 pada seluruh kolom. Hal ini menandakan bahwa data sudah lengkap dan tidak memerlukan proses imputasi atau penanganan data hilang (*missing value handling*), sehingga langsung dapat digunakan untuk proses eksplorasi dan pemodelan.

Terdapat dua tipe data utama dalam dataset ini, yaitu int64 untuk 19 kolom dan float64 untuk 2 kolom. Kolom-kolom bertipe integer (int64) umumnya mewakili fitur-fitur kategorikal biner atau numerik diskrit, seperti blue, dual_sim, four_g, wifi, serta price_range sebagai target variabel. Sementara itu, fitur clock_speed dan m_dep memiliki tipe data float64 yang merepresentasikan nilai desimal, masing-masing untuk kecepatan prosesor dalam GHz dan ketebalan ponsel dalam cm.

turnitin Page 11 of 15 - Integrity Submission Submission ID trn:oid:::1:3479451121

Djtechno : Jurnal Teknologi Informasi
Vol. xxx, No. xxx, Bulan Tahun
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Secara keseluruhan, struktur dataset ini cukup ideal untuk digunakan dalam algoritma klasifikasi seperti *Logistic Regression*, karena tidak memerlukan pembersihan data yang kompleks dan sudah memiliki kombinasi antara fitur numerik dan kategorikal yang relevan terhadap kelas target (price_range). Dengan demikian, data dapat langsung dilanjutkan ke proses pra-pemrosesan dan normalisasi sebelum dilakukan pelatihan model.

3. Statistik Deskriptif Data

Fungsi df.describe() menghasilkan ringkasan statistik seperti nilai rata-rata, standar deviasi, nilai maksimum dan minimum, serta kuartil dari setiap fitur numerik. Tahapan ini penting untuk mengamati sebaran awal data.

```
[5] df.describe()
```

	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	n_cores	...	px_height	px_width	ram	sc_h	sc_w	talk_time	three_g	touch_screen	wifi
count	2000.000000	2000.0000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	...	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000
mean	1238.518500	0.4950	1.522250	0.509500	4.309500	0.521500	32.046500	0.501750	140.249000	4.520500	...	645.108000	1251.515500	2124.213000	12.306500	5.767000	11.011000	0.761500	0.503000	0.507000
std	439.418206	0.5001	0.816004	0.500035	4.341444	0.499662	18.145715	0.288416	35.399655	2.287837	...	443.780811	432.199447	1084.732044	4.213245	4.356398	5.463955	0.426273	0.500116	0.500076
min	501.000000	0.0000	0.500000	0.000000	0.000000	0.000000	2.000000	0.100000	80.000000	1.000000	...	0.000000	500.000000	256.000000	5.000000	0.000000	2.000000	0.000000	0.000000	0.000000
25%	851.750000	0.0000	0.700000	0.000000	1.000000	0.000000	16.000000	0.200000	109.000000	3.000000	...	282.750000	874.750000	1207.500000	9.000000	2.000000	6.000000	1.000000	0.000000	0.000000
50%	1226.000000	0.0000	1.500000	1.000000	3.000000	1.000000	32.000000	0.500000	141.000000	4.000000	...	564.000000	1247.000000	2146.500000	12.000000	5.000000	11.000000	1.000000	1.000000	1.000000
75%	1615.250000	1.0000	2.200000	1.000000	7.000000	1.000000	48.000000	0.800000	170.000000	7.000000	...	947.250000	1633.000000	3064.500000	16.000000	9.000000	16.000000	1.000000	1.000000	1.000000
max	1998.000000	1.0000	3.000000	1.000000	19.000000	1.000000	64.000000	1.000000	200.000000	8.000000	...	1960.000000	1998.000000	3998.000000	19.000000	18.000000	20.000000	1.000000	1.000000	1.000000

8 rows × 21 columns

Gambar 5. Statistik Deskriptif Dataset

Berdasarkan hasil keluaran dari fungsi df.describe(), diketahui bahwa dataset memiliki jumlah data sebanyak 2.000 entri untuk setiap fitur (kolom), tanpa adanya nilai yang hilang. Setiap kolom ditampilkan dengan nilai-nilai statistik deskriptif seperti mean (rata-rata), standard deviation (simpangan baku), min (nilai minimum), max (nilai maksimum), serta kuartil (25%, 50%, dan 75%), yang memberikan gambaran umum terhadap sebaran dan distribusi data.

Fitur battery_power memiliki nilai rata-rata sebesar 1238 mAh dengan nilai minimum 501 dan maksimum 1998, menunjukkan bahwa dataset mencakup ponsel dengan daya baterai yang sangat bervariasi. Fitur ram juga menunjukkan variabilitas tinggi, dengan nilai rata-rata 2.124 MB, minimum 256 MB, dan maksimum mencapai 3.998 MB. Hal ini mengindikasikan bahwa RAM merupakan fitur penting yang dapat digunakan sebagai indikator terhadap harga, yang juga diperkuat oleh perbedaan yang mencolok antara kuartil bawah (25% = 1207.5 MB) dan kuartil atas (75% = 3064.5 MB).

Fitur px_height dan px_width, yang merepresentasikan resolusi layar, juga menunjukkan penyebaran yang luas, dengan px_height mulai dari 0 hingga 1960 piksel dan px_width dari 500 hingga 1998 piksel. Adanya nilai minimum 0 untuk px_height menandakan kemungkinan outlier atau anomali dalam data yang perlu diperhatikan. Demikian pula, fitur clock_speed memiliki nilai rata-rata 1.52 GHz dengan rentang dari 0.5 GHz hingga 3.0 GHz, yang menunjukkan adanya ponsel dengan performa prosesor rendah hingga tinggi.

turnitin Page 11 of 15 - Integrity Submission Submission ID trn:oid:::1:3479451121

Djtechno : Jurnal Teknologi Informasi
Vol. xxx, No. xxx, Bulan Tahun
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Fitur-fitur biner seperti blue, dual_sim, four_g, three_g, touch_screen, dan wifi hanya memiliki nilai minimum 0 dan maksimum 1, yang menunjukkan bahwa fitur-fitur ini merepresentasikan keberadaan (1) atau ketiadaan (0) dari teknologi tersebut pada perangkat.

b. Hasil

Setelah model *Logistic Regression* selesai dibangun dan dievaluasi menggunakan data latih dan uji internal, langkah selanjutnya adalah melakukan pengujian terhadap data eksternal atau data uji tanpa label. Data ini berasal dari file CSV bernama test.csv yang diunggah dan dibaca menggunakan fungsi pd.read_csv() dari pustaka pandas.

```
[49] df3=pd.read_csv('/content/test.csv')
```

Dataset df3 yang dimuat berisi data spesifikasi smartphone tanpa kolom price_range. Sebelum digunakan dalam proses prediksi, data ini terlebih dahulu diproses agar memiliki skala dan struktur yang sama dengan data pelatihan.

	id	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	...	pc	px_height	px_width	ram	sc_h	sc_w	talk_time	three_g	touch_screen	wifi
0	1	1043	1	1.8	1	14	0	5	0.1	193	...	16	226	1412	3476	12	7	2	0	1	0
1	2	841	1	0.5	1	4	1	61	0.8	191	...	12	746	857	3895	6	0	7	1	0	0
2	3	1807	1	2.8	0	1	0	27	0.9	186	...	4	1270	1366	2396	17	10	10	0	1	1
3	4	1546	0	0.5	1	18	1	25	0.5	96	...	20	295	1752	3893	10	0	7	1	1	0
4	5	1434	0	1.4	0	11	1	49	0.5	108	...	18	749	810	1773	15	8	7	1	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	996	1700	1	1.9	0	0	1	54	0.5	170	...	17	644	913	2121	14	8	15	1	1	0
996	997	609	0	1.8	1	0	0	13	0.9	186	...	2	1152	1632	1933	8	1	19	0	1	1
997	998	1185	0	1.4	0	1	1	8	0.5	80	...	12	477	825	1223	5	0	14	1	0	0
998	999	1533	1	0.5	1	0	0	50	0.4	171	...	12	38	832	2509	15	11	6	0	1	0
999	1000	1270	1	0.5	0	4	1	35	0.1	140	...	19	457	608	2828	9	2	3	1	0	1

1000 rows × 21 columns

Gambar 6. Tampilan Dataset Test

Proses pertama yang dilakukan adalah normalisasi fitur numerik menggunakan metode Min-Max Scaling, sama seperti pada tahap preprocessing sebelumnya. Fitur-fitur yang dinormalisasi antara lain fc, int_memory, mobile_wt, n_cores, pc, px_height, ram, sc_h, sc_w, talk_time, dan battery_power. Proses normalisasi ini penting agar prediksi model tetap akurat dan tidak bias terhadap perbedaan skala.

```
[51] cols_scale=df3[['fc','int_memory','mobile_wt','n_cores','pc','px_height','ram','sc_h' ,'sc_w' ,'talk_time', 'px_height']]

[52] for i in cols_scale:
         df3[i]=df3[i]=minmax.fit_transform(df3[[i]])
```

Setelah proses scaling selesai, dilakukan pemilahan data uji ke dalam variabel x_pred, yang berisi seluruh fitur input (kecuali price_range, karena belum tersedia). Fitur-fitur ini mencerminkan spesifikasi smartphone secara lengkap, seperti daya baterai, RAM, ukuran layar, dan konektivitas.

Djtechno : Jurnal Teknologi Informasi
Vol. xxx, No. xxx, Bulan Tahun
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

	id	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	...	pc	px_height	px_width	ram	sc_h	sc_w	talk_time	three_g	touch_screen	wifi
0	1	0.362241	1	1.8	1	0.736842	0	0.048387	0.1	0.941667	...	0.80	0.118511	1412	0.862319	0.500000	0.388889	0.000000	0	1	0
1	2	0.227485	1	0.5	1	0.210526	1	0.951613	0.8	0.925000	...	0.60	0.391190	857	0.974772	0.071429	0.000000	0.277778	1	0	0
2	3	0.871915	1	2.8	0	0.052632	0	0.403226	0.9	0.883333	...	0.20	0.665967	1366	0.572464	0.857143	0.555556	0.444444	0	1	1
3	4	0.897799	0	0.5	1	0.947368	1	0.370968	0.5	0.133333	...	1.00	0.154693	1752	0.974235	0.357143	0.000000	0.277778	1	1	0
4	5	0.623082	0	1.4	0	0.578947	1	0.758065	0.5	0.233333	...	0.90	0.392764	810	0.405260	0.714286	0.444444	0.277778	1	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	996	0.800534	1	1.9	0	0.000000	1	0.838710	0.5	0.750000	...	0.85	0.337703	913	0.498658	0.642857	0.444444	0.722222	1	1	0
996	997	0.072715	0	1.8	1	0.000000	0	0.177419	0.9	0.883333	...	0.10	0.604090	1632	0.448202	0.214286	0.055556	0.944444	0	1	1
997	998	0.456971	0	1.4	0	0.052632	1	0.096774	0.5	0.000000	...	0.60	0.250131	825	0.257649	0.000000	0.000000	0.666667	1	0	0
998	999	0.689126	1	0.5	1	0.000000	0	0.774194	0.4	0.758333	...	0.60	0.019927	832	0.602791	0.714286	0.611111	0.222222	0	1	0
999	1000	0.513676	1	0.5	0	0.210526	1	0.532258	0.1	0.500000	...	0.95	0.239643	608	0.688406	0.285714	0.111111	0.055556	1	0	1

1000 rows × 21 columns

Gambar 7. Tampilan Dataset Test Normalisasi

Model yang telah dilatih sebelumnya digunakan untuk melakukan prediksi terhadap x_pred dengan fungsi model.predict(). Hasil prediksi disimpan dalam variabel y2_pred, yang berisi **kelas harga smartphone** hasil klasifikasi berdasarkan spesifikasi teknis. Prediksi tersebut kemudian ditambahkan ke dalam dataset df3 sebagai kolom baru bernama price_range, sehingga setiap baris data dalam file uji kini memiliki nilai kelas harga yang diprediksi.

Sebagai tambahan, dilakukan pengujian akurasi prediksi model terhadap data yang sama menggunakan fungsi model.score(x_pred, y2_pred), untuk memastikan konsistensi model pada data uji.

```
model.score(x_pred,y2_pred)
```

```
1.0
```

Meskipun nilai ini tidak mewakili akurasi riil (karena tidak ada ground truth), hasil ini dapat digunakan sebagai pengukuran **self-consistency** model terhadap prediksi yang dihasilkannya.

```
[59] df3['price_range']=y2_pred
```

```
[60] df3
```

	id	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	...	px_height	px_width	ram	sc_h	sc_w	talk_time	three_g	touch_screen	wifi	price_range
0	1	0.362241	1	1.8	1	0.736842	0	0.048387	0.1	0.941667	...	0.118511	1412	0.862319	0.500000	0.388889	0.000000	0	1	0	3
1	2	0.227485	1	0.5	1	0.210526	1	0.951613	0.8	0.925000	...	0.391190	857	0.974772	0.071429	0.000000	0.277778	1	0	0	3
2	3	0.871915	1	2.8	0	0.052632	0	0.403226	0.9	0.883333	...	0.665967	1366	0.572464	0.857143	0.555556	0.444444	0	1	1	3
3	4	0.897799	0	0.5	1	0.947368	1	0.370968	0.5	0.133333	...	0.154693	1752	0.974235	0.357143	0.000000	0.277778	1	1	0	3
4	5	0.623082	0	1.4	0	0.578947	1	0.758065	0.5	0.233333	...	0.392764	810	0.405260	0.714286	0.444444	0.277778	1	0	1	3
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	996	0.800534	1	1.9	0	0.000000	1	0.838710	0.5	0.750000	...	0.337703	913	0.498658	0.642857	0.444444	0.722222	1	1	0	3
996	997	0.072715	0	1.8	1	0.000000	0	0.177419	0.9	0.883333	...	0.604090	1632	0.448202	0.214286	0.055556	0.944444	0	1	1	3
997	998	0.456971	0	1.4	0	0.052632	1	0.096774	0.5	0.000000	...	0.250131	825	0.257649	0.000000	0.000000	0.666667	1	0	0	3
998	999	0.689126	1	0.5	1	0.000000	0	0.774194	0.4	0.758333	...	0.019927	832	0.602791	0.714286	0.611111	0.222222	0	1	0	3
999	1000	0.513676	1	0.5	0	0.210526	1	0.532258	0.1	0.500000	...	0.239643	608	0.688406	0.285714	0.111111	0.055556	1	0	1	3

1000 rows × 22 columns

Gambar 8. Tampilan Hasil Klasifikasi Price_range pada data test

Dengan tahapan ini, penelitian menunjukkan bahwa model yang dibangun tidak hanya efektif pada data pelatihan dan pengujian internal, tetapi juga **siap diterapkan pada data eksternal** dalam situasi nyata untuk memprediksi kelas harga smartphone secara otomatis.

Djtechno : Jurnal Teknologi Informasi
Vol. xxx, No. xxx, Bulan Tahun
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai penerapan algoritma Logistic Regression dalam klasifikasi harga smartphone berdasarkan spesifikasi teknisnya, dapat disimpulkan beberapa hal sebagai berikut:

1. Model Logistic Regression terbukti mampu digunakan untuk mengklasifikasikan harga smartphone ke dalam empat kelas (price_range) berdasarkan fitur-fitur seperti battery_power, ram, px_height, clock_speed, dan fitur lainnya dengan tingkat akurasi yang cukup baik.

2. Proses eksplorasi data menunjukkan bahwa fitur-fitur seperti RAM dan resolusi layar (px_height, px_width) memiliki korelasi kuat terhadap kelas harga, yang selanjutnya menjadi penentu utama dalam pelatihan model klasifikasi.

3. Melalui tahapan preprocessing, khususnya normalisasi menggunakan MinMaxScaler, model dapat memproses data dengan skala yang seragam, sehingga meningkatkan kestabilan hasil prediksi.

4. Hasil evaluasi menggunakan classification report dan confusion matrix menunjukkan bahwa model menghasilkan performa klasifikasi yang cukup seimbang antar kelas, meskipun masih ditemukan beberapa kesalahan klasifikasi, terutama pada kelas harga yang berdekatan (misalnya kelas 1 dan kelas 2).

5. Pengujian terhadap data CSV eksternal tanpa label menunjukkan bahwa model Logistic Regression dapat digunakan untuk melakukan prediksi harga secara otomatis berdasarkan spesifikasi, yang mendukung aplikasi praktis di dunia industri atau e-commerce

Djtechno : Jurnal Teknologi Informasi
Vol. xxx, No. xxx, Bulan Tahun
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Universitas Potensi Utama yang telah banyak memberikan masukkan dan saran dalam penyelesaian skripsi ini.

REFERENCES

[1] Ardelia, D. N., Arifin, H. D., Daniswara, S., & Sari, A. P. (2024) Klasifikasi Harga Ponsel Menggunakan Algoritma Logistic Regression.

[2] Saputri, I., Arsi, P., & Isnaini, K. N. (2025). EFFECTIVENESS HYPERPARAMETER TUNING ON RANDOM FOREST, LINEAR DISCRIMINANT ANALYSIS, LOGISTIC REGRESSION AND NAÏVE BAYES ALGORITHMS FOR DETECTING DOS NETWORK ATTACKS. Jurnal Teknik Informatika (Jutif), 6(1), 87-104.

[3] Subhiksha, S., Thota, S., & Sangeetha, J. (2020). Prediction of phone prices using machine learning techniques. In Data Engineering and Communication Technology: Proceedings of 3rd ICDECT-2K19 (pp. 781-789). Springer Singapore.

[4] Ash, S., & Surya, A. (2022). Analisis sentimen masyarakat terhadap kebijakan vaksinasi covid-19 pada media sosial twitter menggunakan metode logistic regression. Analisis Sentimen Masyarakat Terhadap Kebijakan Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Metode Logistic Regression, 3(2), 99-106.

[5] Sunariya, N., Singh, A., Alam, M., & Gaur, V. (2024). Classification of Mobile Price Using Machine Learning. In SCI (pp. 55-66).

[6] Liang, Q. (2024). Mobile phone price prediction: A comparative study among four models. Applied and Computational Engineering, 48, 212-218.

[7] Ercan, S. İ. A., & ŞİMŞEK, M. (2023). Mobile Phone Price Classification Using Machine Learning. International Journal of Advanced Natural Sciences and Engineering Researches, 7(4), 458-462.

[8] Lashari, S. A., Khan, M. M., Khan, A., Salahuddin, S., & Ata, M. N. (2024). Comparative Evaluation of Machine Learning Models for Mobile Phone Price Prediction: Assessing Accuracy, Robustness, and Generalization Performance. Journal of Informatics and Web Engineering, 3(3), 147-163.

[9] Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. Sensors, 21(11), 3758.

[10] Raschka, S., Liu, Y. H., & Mirjalili, V. (2022). Machine Learning with PyTorch and Scikit- Learn: Develop machine learning and deep learning models with Python. Packt Publishing Ltd.

[11] Lashari, S. A., Khan, M. M., Khan, A., Salahuddin, S., & Ata, M. N. (2024). Comparative Evaluation of Machine Learning Models for Mobile Phone Price Prediction: Assessing Accuracy, Robustness, and Generalization Performance. Journal of Informatics and Web Engineering, 3(3), 147-163.

[12] Lu, J. (2025). Prediction on Mobile Phone Price Rangeusing Classification Models. Science and Technology of Engineering, Chemistry and Environmental Protection, 1(3).

[13] Sunariya, N., Singh, A., Alam, M., & Gaur, V. (2024). Classification of Mobile Price Using Machine Learning. In SCI (pp. 55-66).

[14] Saputra, P. S., TI, S., Kom, M., & Sudiatmika, I. P. G. A. (2024). ANALISIS PREDIKSI HARGA SMARTPHONE TAHUN 2023 MENGGUNAKAN MODEL RANDOM FOREST REGRESSION BERDASARKAN FITUR-FITUR SPESIFIKASI TEKNIS. KOMTEKS, 3(2).

[15] Geng, Y., Li, Q., Yang, G., & Qiu, W. (2024). Logistic regression. In Practical Machine Learning Illustrated with KNIME (pp. 99-132). Singapore: Springer Nature Singapore.