

Analisis Data Segmentasi Produk Smartphone Berbasis Machine Learning

Novia Lestari^{✉1}, M. Rafly Syahrul Fathan²

Sistem Informasi, Universitas Islam Negeri Imam Bonjol Padang, Indonesia^{1,2}
email: noviaa.lestari@uinib.ac.id¹, prapthorespect@gmail.com²

Received 02 September 2024, Accepted 17 Oktober 2024, Published 31 Maret 2025

Abstrak

Di era digital yang berkembang pesat, industri smartphone telah mengalami pertumbuhan yang signifikan dan menjadi salah satu sektor yang paling kompetitif. Perusahaan smartphone terus berinovasi untuk memenuhi beragam kebutuhan dan preferensi konsumen dengan melahirkan berbagai jenis smartphone dengan spesifikasi yang beragam. Hal ini menyebabkan konsumen kesulitan dalam menentukan spesifikasi smartphone yang cocok yang sesuai dengan kebutuhan dan budget mereka. Bagi perusahaan, memahami perilaku konsumen dan mengidentifikasi tren pembelian yang mendasarinya juga merupakan tantangan yang kompleks. Penelitian Analisis Data Segmentasi Produk Smartphone Berbasis Machine Learning ini dirancang untuk mengkategorikan produk smartphone ke dalam empat segmen utama berdasarkan berbagai atribut dan perilaku pembelian konsumen, sehingga konsumen memiliki gambaran tentang spesifikasi produk dan harga smartphone yang sesuai dengan kebutuhan dan kemampuan mereka. Selain itu, hasil analisis segmentasi produk smartphone ini juga bisa menjadi alat yang berharga untuk membantu perusahaan dalam mengembangkan strategi pemasaran yang lebih efektif untuk menghasilkan produk yang lebih relevan untuk pasar mereka. Aplikasi ini menggunakan data penjualan smartphone yang bersumber dari kaggle dan menerapkan kerangka kerja CRISP-DM (*Cross Industry Standard Process for Data Mining*) untuk mengklasifikasikan produk smartphone.

Kata Kunci: Analisis; Data Mining; Machine Learning

Abstract

In the rapidly growing digital era, the smartphone industry has experienced significant growth and become one of the most competitive industries. Smartphone companies continue to innovate to meet the diverse needs and preferences of consumers by creating different types of smartphones with varying specifications. It's making consumers difficult to determine the appropriate mobile phone specifications that meet their needs and budget. For companies, understanding consumer behavior and identifying underlying buying trends is also a complex challenge. Therefore, smartphone product segmentation analysis can be a valuable tool to help companies develop more effective marketing strategies and more relevant products for their market. The Machine Learning Based Smartphone Product Segmentation Data Analyze application is designed to categorize smartphone products into four main segments based on various attributes and consumer purchasing behavior. This application uses mobile phone sales data from Kaggle and applies the Cross Industry Standard Process for Data Mining (CRISP-DM) framework to classify smartphone products.

Kata Kunci: Analisis; Data Mining; Machine Learning

✉ Corresponding author

PENDAHULUAN

Di Era Revolusi Industri 4.0 saat ini, telepon pintar atau smartphone telah menjadi bagian penting dari kehidupan masyarakat. Hal ini terbukti dengan perkembangan penggunaan smartphone di Indonesia yang terus meningkat setiap tahunnya [1]. Menurut data Badan Pusat Statistik (BPS), 67,88% penduduk Indonesia yang berusia 5 tahun ke atas sudah memiliki ponsel atau *handphone* pada 2022. Persentase tersebut meningkat dibanding 2021 yang masih 65,87%, sekaligus menjadi rekor tertinggi dalam sedekade terakhir [2]. Hal ini membuktikan bahwa smartphone telah menjadi industri yang penting secara ekonomi karena menghasilkan keuntungan yang besar dalam berbagai aspek kehidupan, baik untuk keperluan pribadi maupun bisnis [3]. Selain berfungsi sebagai alat komunikasi, smartphone juga berperan sebagai sumber informasi, hiburan, dan penyimpanan berbagai data penting untuk pengelolaan keuangan, pendidikan, dan kebutuhan pribadi lainnya. Smartphone juga telah membawa revolusi dalam industri seperti pertanian, kehutanan, dan pemantauan kesehatan dengan memungkinkan pengumpulan, analisis, dan komunikasi data yang lebih efisien [4].

Dengan keterkaitan antara fungsi smartphone dan kehidupan sehari-hari tersebut, industri smartphone terus berkembang dan persaingan antar produsen semakin ketat yang mengakibatkan munculnya banyak merk smartphone dengan menghadirkan berbagai macam model dan spesifikasi produk yang semakin lengkap dan canggih [5]. Para produsen menawarkan pengalaman dan fitur yang unik mulai dari aspek produktivitas, inovasi, hingga kenyamanan. Mereka hadir dengan produk unggulan mereka sendiri dengan melakukan inovasi secara konsisten untuk tetap bersaing di pasar yang dinamis ini. Hal ini menyebabkan konsumen kesulitan dalam menentukan spesifikasi smartphone yang cocok yang sesuai dengan kebutuhan dan budget mereka. Bagi perusahaan, memahami perilaku konsumen dan mengidentifikasi tren pembelian yang mendasarinya juga merupakan tantangan yang kompleks [6]. Oleh karena itu, analisis segmentasi produk menjadi alat yang berharga untuk membantu perusahaan dalam mengembangkan strategi pemasaran yang lebih efektif dan produk yang lebih relevan untuk pasar mereka [7].

Aplikasi Analisis Data Segmentasi Produk Smartphone Berbasis Machine Learning menggunakan data penjualan *handphone* yang bersumber dari kaggle yang dirancang untuk mengkategorikan produk smartphone ke dalam empat segmen utama berdasarkan berbagai atribut dan perilaku pembelian konsumen. Hasil analisis ini akan memberikan panduan yang berharga bagi konsumen dalam memilih smartphone yang sesuai dengan kebutuhan dan budget mereka. Sedangkan bagi perusahaan, hasil analisis ini dapat berkontribusi dalam membuat keputusan strategis yang lebih akurat terkait pangsa pasar penjualan produk smartphone sebagai upaya meningkatkan daya saing perusahaan di pasar yang dinamis dan kompetitif.

METODOLOGI

Penelitian ini menerapkan kerangka kerja CRISP-DM (*Cross Industry Standard Process for Data Mining*) dengan algoritma *machine learning* berupa K-Means *clustering* untuk melakukan segmentasi produk pada data di atas. Kerangka kerja CRISP-DM adalah kerangka kerja data mining yang menerapkan enam fase dalam metodenya untuk mengidentifikasi input dan output dalam suatu proses. Keenam fase tersebut adalah *business understanding*, *data understanding*, *data preparation*, *modelling*, *deployment*, dan evaluasi [8] sebagai berikut :

1. Business Understanding

Tujuan tahap ini adalah untuk menilai kondisi keadaan bisnis untuk memperoleh pemahaman tentang sumber daya yang dibutuhkan. Menetapkan target data mining dengan melakukan identifikasi dan analisis terhadap tantangan yang dihadapi, tujuan bisnis yang ingin dicapai, serta menyusun rencana pengembangan proyek machine learning.

2. Data Understanding

Penulis melakukan pengumpulan data yang diperlukan, mengevaluasi dan mengeksplorasi data, serta memastikan kualitas data. Melalui tahap ini, didapatkan kesimpulan terkait data yang akan digunakan dan strategi yang harus diambil untuk mengatasi data tersebut. Pengumpulan data dilakukan dengan melalui kaggle.com. Penulis juga melakukan *exploratory data analysis* (EDA) yang merupakan proses dimana representasi dalam bentuk numerik, tabel ataupun grafik dibuat untuk memberi rangkuman serta menerangkan aspek dan fitur utama dari suatu data.

3. Data Preparation

Data yang sudah dieksplorasi pada tahap sebelumnya harus dilakukan *preprocessing*, dibersihkan dan dinilai dimensionalitas datanya. Tahap pembersihan yaitu dengan menghapus atau mengganti *null value* atau nilai data yang *error*. Selanjutnya, dengan menilai dimensionalitas data, dapat ditentukan atribut data yang perlu dan tidak perlu dihilangkan.

4. Data Modelling

Pengembangan *data model* dibantu dengan *tools* Jupyter Notebook.

5. Setelah tahap *modeling* selesai, dilakukan *deployment* melalui sebuah *web app* yang dikembangkan menggunakan *microservice* Flask dengan bahasa pemrograman utama Python.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan dataset *mobile_phone_prize* yang diambil dari kaggle.com yang terdiri dari 407 data dengan 8 variabel, yaitu : Brand, model, Storage, RAM, Screen size (inches), Camera (MP), Battery Capacity (mAh), Price (\$). Data yang sudah dieksplorasi pada tahap sebelumnya dibersihkan dan dinilai dimensionalitas datanya melalui tahapan *preprocessing* data sebagai berikut :

1. Cleaning data

Pada tahapan *cleaning* ini digunakan digunakan untuk menyusun dan memperbaiki dataset yang tidak tepat atau memiliki format yang salah atau tidak beraturan. Proses *cleaning* dalam penelitian ini yaitu menghapus beberapa data yang tidak valid serta mengubah bentuk data frame menjadi lebih sederhana sehingga mudah untuk dilakukan

analisis. Tahapan awal dalam proses cleaning ini yaitu menghilangkan baris dengan nilai yang hilang. Kemudian menghapus spasi disekitar nama kolom dan mengubah format data untuk variable "price". Format data yang akan diubah yaitu menghilangkan simbol dollar (\$) yang berada dalam data set dan mengubah tipe datanya menjadi numeric. Kemudian menghilangkan GB pada variable "data storage" dan "ram" dan mengubah tipe data menjadi numeric. Selanjutnya menghilangkan MP pada variable "camera" serta mengkonversi setiap bagian menjadi float dan menghitung jumlah MP dalam data camera agar dapat memudahkan dalam melakukan analisis data. Berikut tampilan dataset setelah dilakukan proses cleaning data seperti terlihat pada Gambar 1 berikut :

	Brand	Model	Storage	RAM	Screen_Size_(inches)	Camera_(MP)	Battery_Capacity_(mAh)	Price_(\$)
0	Apple	iPhone 13 Pro	128	6	6.10	36.0	3095	999
1	Samsung	Galaxy S21 Ultra	256	12	6.80	140.0	5000	1199
2	OnePlus	9 Pro	128	8	6.70	108.0	4500	899
3	Xiaomi	Redmi Note 10 Pro	128	6	6.67	79.0	5020	279
4	Google	Pixel 6	128	8	6.40	62.2	4614	799
...	...	...	...	...	...	...	...	...
402	Samsung	Galaxy Note20 5G	128	8	6.70	88.0	4300	1049
403	Xiaomi	Mi 10 Lite 5G	128	6	6.57	60.0	4160	349
404	Apple	iPhone 12 Pro Max	128	6	6.70	36.0	3687	1099
405	Oppo	Reno3	128	8	6.40	71.0	4025	429
406	Samsung	Galaxy S10 Lite	128	6	6.70	65.0	4500	649

407 rows x 8 columns

Gambar 1. Data Hasil Pre Processing

Tampilan diatas merupakan tampilan dataframe yang telah dilakukan proses *cleaning*. Dari tampilan dataset tersebut sudah memiliki banyak perubahan seperti tampilan dataframe yang telah menjadi bentuk yang lebih sederhana, serta sudah tidak ada lagi data set yang bernilai kosong.

2. Data Transformasi

Dalam tahapan data transformasi ini kita akan mengkategorikan price atau kategori harga menjadi 4 kategori yaitu : 1. Bronze, 2.Silver, 3.Gold, dan 4. Platinum. Series yang memiliki label untuk setiap baris berdasarkan ke dalam kuartil mana nilai Price (\$) tersebut jatuh. Misalnya, jika harga sebuah ponsel berada di kuartil terendah, maka labelnya akan menjadi 'Bronze', dan jika berada di kuartil tertinggi, labelnya akan menjadi 'Platinum'. Berikut hasil pengelompokan berdasarkan kategori harga seperti terlihat pada Tabel 1 :

Tabel 1. Pengelompokan Kategori Harga

Result	Count
Platinum	95
Gold	92
Silver	101
Bronze	119

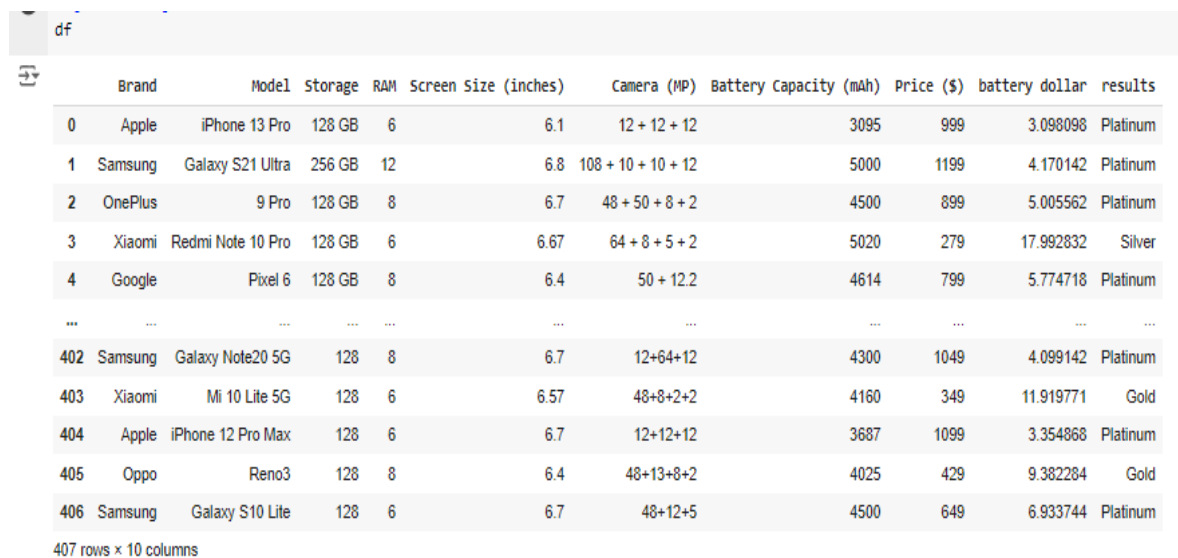
Selanjutnya mengelompokkan data berdasarkan kolom 'Brand' dan 'results' dan menghitung jumlah entri (baris) dalam setiap grup, kemudian mengubah hasil

pengelompokan menjadi format tabel dengan kolom yang terpisah untuk setiap kategori harga (Bronze, Silver, Gold, Platinum) seperti yang terlihat pada Tabel 2 berikut:

Tabel 2. Pengelompokkan Brand dan Result

Brand	Result			
	Bronze	Silver	Gold	Platinum
Apple	0	0	8	22
Asus	0	0	0	4
Blackberry	0	0	2	1
Huawei	1	0	3	8
LG	0	0	2	1
Motorola	8	10	4	1
Nokia	17	5	4	2
Oppo	14	17	14	11
Realme	26	12	5	0
Samsung	17	19	22	21
Sony	0	0	0	1
Vivo	11	13	8	3
Xiaomi	25	22	15	5

Setelah data transformasi dilakukan, dapat diketahui variable target (result) dari dataset yang ada seperti yang terlihat pada Gambar 2 berikut:



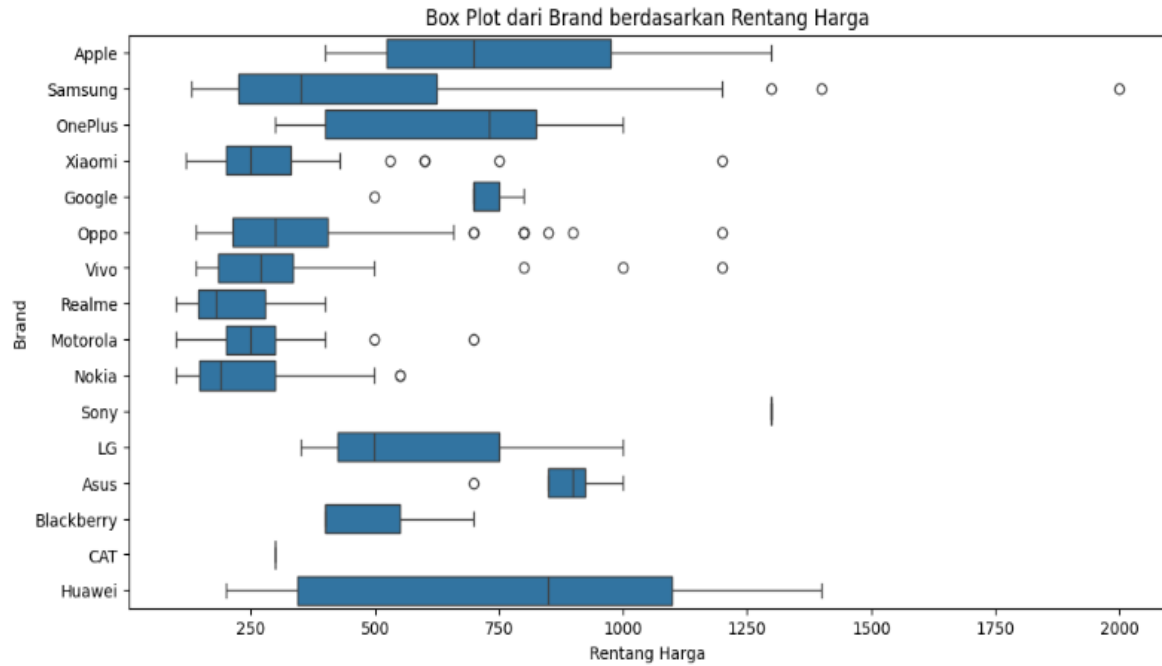
	Brand	Model	Storage	RAM	Screen Size (inches)	Camera (MP)	Battery Capacity (mAh)	Price (\$)	battery dollar	results
0	Apple	iPhone 13 Pro	128 GB	6	6.1	12 + 12 + 12	3095	999	3.098098	Platinum
1	Samsung	Galaxy S21 Ultra	256 GB	12	6.8	108 + 10 + 10 + 12	5000	1199	4.170142	Platinum
2	OnePlus	9 Pro	128 GB	8	6.7	48 + 50 + 8 + 2	4500	899	5.005562	Platinum
3	Xiaomi	Redmi Note 10 Pro	128 GB	6	6.67	64 + 8 + 5 + 2	5020	279	17.992832	Silver
4	Google	Pixel 6	128 GB	8	6.4	50 + 12.2	4614	799	5.774718	Platinum
...	...	...	...	...	...	...	...	...	...	...
402	Samsung	Galaxy Note20 5G	128	8	6.7	12+64+12	4300	1049	4.099142	Platinum
403	Xiaomi	Mi 10 Lite 5G	128	6	6.57	48+8+2+2	4160	349	11.919771	Gold
404	Apple	iPhone 12 Pro Max	128	6	6.7	12+12+12	3687	1099	3.354868	Platinum
405	Oppo	Reno3	128	8	6.4	48+13+8+2	4025	429	9.382284	Gold
406	Samsung	Galaxy S10 Lite	128	6	6.7	48+12+5	4500	649	6.933744	Platinum

407 rows x 10 columns

Gambar 2. Data Hasil Pengelompokkan

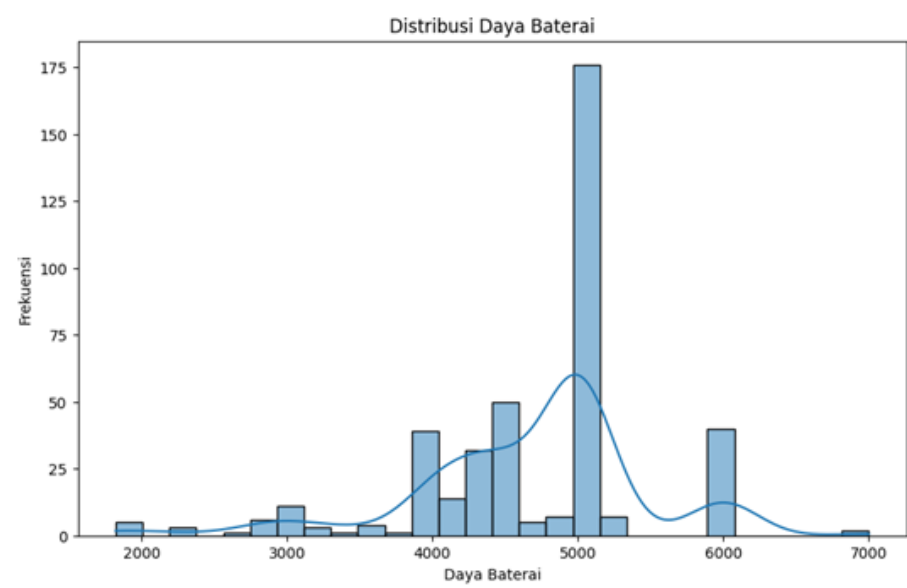
Berdasarkan hasil pengelompokkan di atas, *brand smartphone* yang menempati peringkat teratas dilihat dari result (kategori harga) platinum adalah Apple dengan model Iphone 13 Pro dengan *Storage* 128 GB, *RAM* 6 GB, *Screen Size* 6.1, *camera* 12 MP, dan *battery capacity* hanya 3095 mAh, dimana spesifikasi nya malah lebih rendah dibandingkan pesaingnya Samsung Galaxy S21 Ultra dengan spesifikasi *storage* 256 GB, *RAM* 12 GB, *screen size* 6.8, dan *battery capacity* yang tinggi yaitu 5000 mAh.

Dari data hasil pengelompokkan di atas, dapat dilihat juga visualisasi dalam bentuk boxplot dari brand berdasarkan rentang harga sesuai Gambar 3 berikut:



Berdasarkan Gambar 3 di atas, dari rentang harga keseluruhan, brand Apple memang menempatkan seluruh model/jenis produknya dikisaran harga yang seimbang, yaitu rata-rata 300 – 1300\$, berbeda dengan pesaingnya Samsung, yang kisaran harga produknya bervariasi, mulai dari harga terendah 200\$ sampai harga tertinggi 2000\$. Selain itu, Brand Oppo juga memiliki rentang harga yang bervariasi untuk masing-masing jenisnya, mulai dari harga terendah 200\$ sampai harga tertinggi 1250\$.

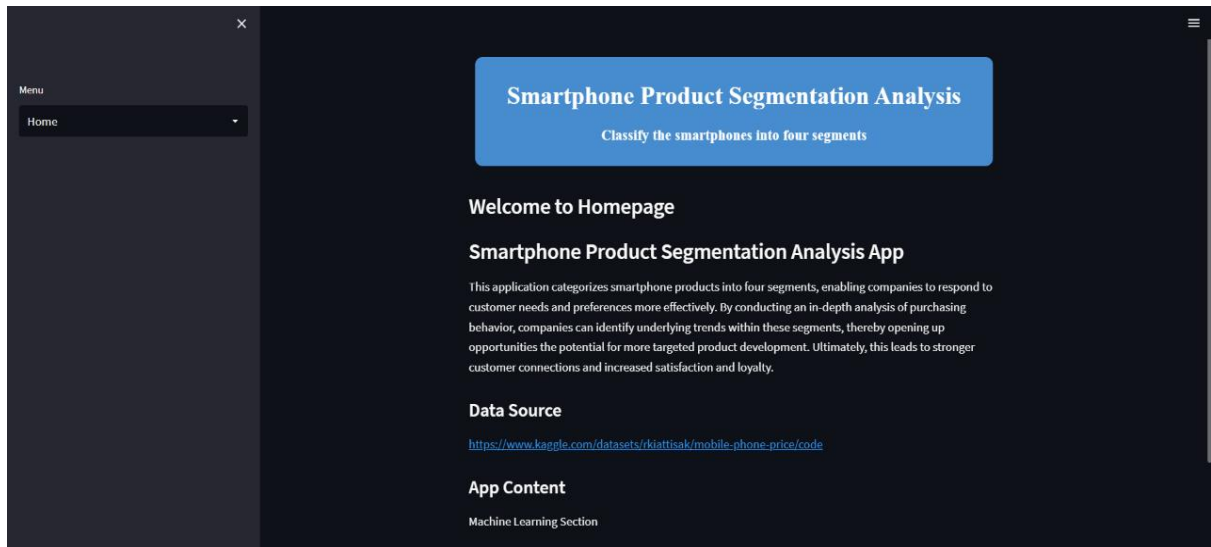
Visualisasi lain yang dihasilkan dari analisis segmentasi produk smartphone berupa distribusi daya baterai seperti terlihat pada Gambar 4 berikut :



Gambar 4. Visualisasi Data Distribusi Daya Baterai

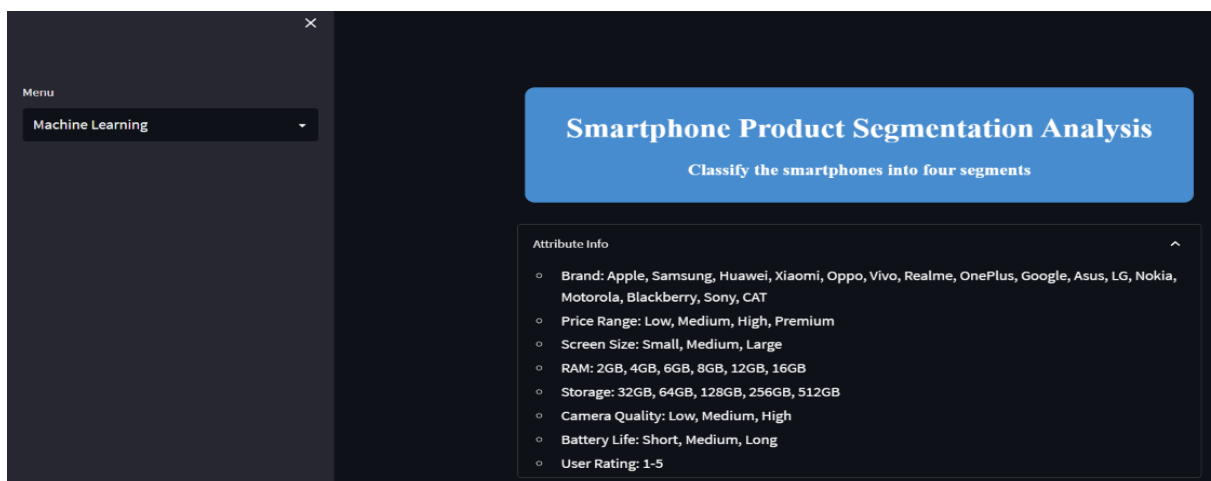
Berdasarkan Gambar 4 di atas, distribusi daya baterai produk smartphone saat ini berkisar antara 2000 hingga 7000 mAh, dengan daya baterai paling banyak digunakan yaitu 5000 mAh.

Setelah dilakukan pengembangan data model dibantu dengan tools Jupyter Notebook, maka dalam penelitian ini, Streamlit digunakan sebagai platform untuk mendeploy model analisis data segmentasi produk smartphone yang telah dikembangkan sebelumnya. Dengan menggunakan Streamlit, model dapat diakses melalui antarmuka web yang interaktif, memungkinkan pengguna untuk melakukan analisis data segmentasi produk smartphone berdasarkan input variabel yang disediakan. Pendekatan ini memudahkan dalam memvisualisasikan hasil prediksi dan mempercepat proses pengujian serta validasi model di lingkungan yang lebih praktis dan user-friendly. Untuk hasil deploy model dalam penelitian ini dapat dilihat pada Gambar 5 berikut ini:



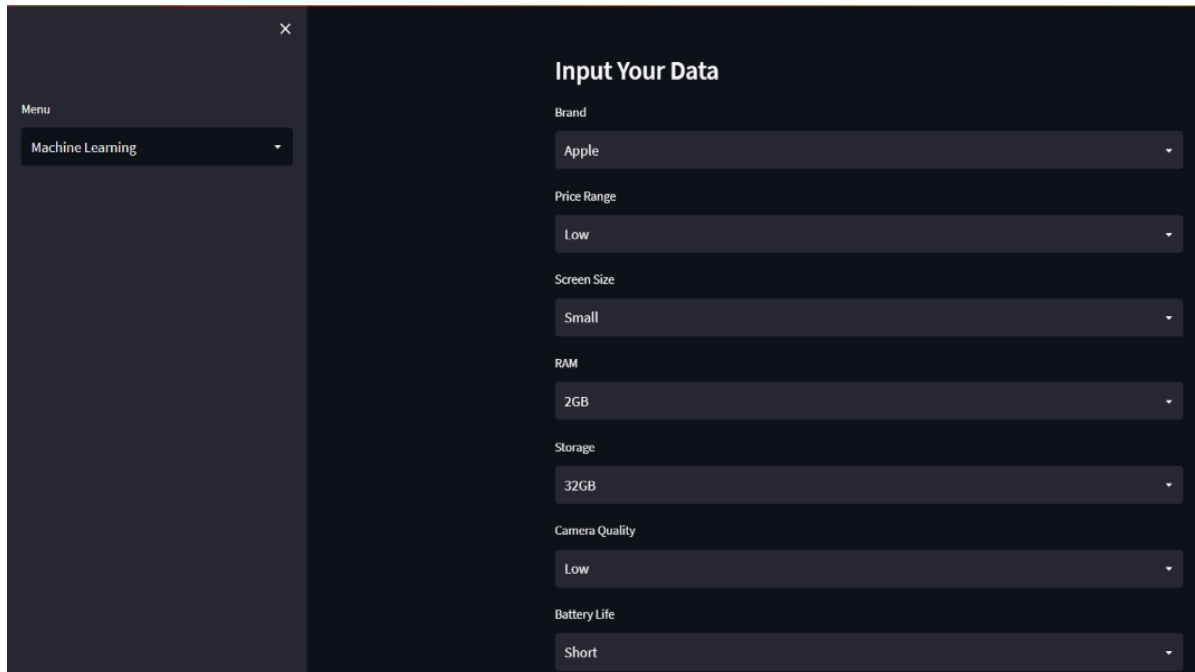
Gambar 5. Tampilan Menu Home

Pada menu home ini, terdapat deskripsi tentang aplikasi analisis data segmentasi produk smartphone. Kemudian pada menu machine learning, terdapat variable-variabel apa saja yang digunakan dalam menjalankan aplikasi untuk segmentasi produk smartphone seperti yang terlihat pada Gambar 6 berikut:



Gambar 6. Tampilan Halaman Machine Learning

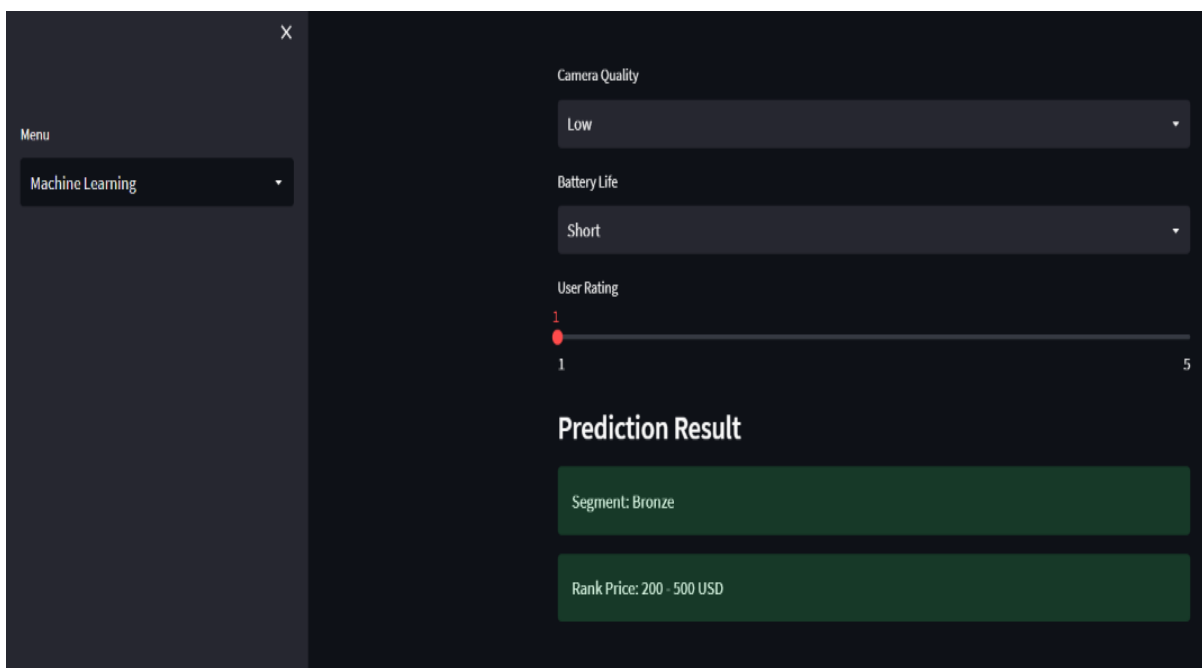
Variabel-variabel tersebut (Brand, Price Range, Screen Size, RAM, Storage, Camera Quality, Battery Life) di inputkan untuk mendapatkan segmentasi produk smartphone berdasarkan spesifikasi yang ada seperti terlihat pada Gambar 7 berikut :



The screenshot shows a web application interface for data input. On the left is a sidebar with a 'Menu' section containing a dropdown for 'Machine Learning'. The main area is titled 'Input Your Data' and contains several form fields, each with a label and a dropdown menu: 'Brand' (Apple), 'Price Range' (Low), 'Screen Size' (Small), 'RAM' (2GB), 'Storage' (32GB), 'Camera Quality' (Low), and 'Battery Life' (Short). The interface is dark-themed with light-colored text and dropdown options.

Gambar 7. Tampilan Halaman Input Data

Setelah semua permintaan inputan telah terisi, akan di tampilkan hasil prediksi segmentasi produk smartphone seperti terlihat pada Gambar 8 berikut ini :



The screenshot displays the 'Prediction Result' section of the application. It shows the 'Camera Quality' dropdown set to 'Low' and the 'Battery Life' dropdown set to 'Short'. Below these is a 'User Rating' slider with a red dot at the '1' mark, ranging from 1 to 5. The results are shown in two green boxes: 'Segment: Bronze' and 'Rank Price: 200 - 500 USD'. The sidebar menu on the left remains the same as in Gambar 7.

Gambar 8. Tampilan Halaman Prediction Result

Berdasarkan Gambar 8 di atas, Prediction result yang dihasilkan berupa segment dan rentang harga (*rank price*) yang disesuaikan dengan pengelompokkan produk smartphone sesuai kriteria yang telah diinputkan.

SIMPULAN

Aplikasi Analisis Data Segmentasi Produk Smartphone Berbasis Machine Learning yang menggunakan data penjualan handphone yang bersumber dari kaggle dapat digunakan untuk mengkategorikan produk smartphone ke dalam empat segmen utama berdasarkan atribut spesifikasi, seperti : Brand, Price Range, Screen Size, RAM, Storage, Camera Quality, Battery Life dan variable rank price (harga). Hasil analisis ini memberikan panduan yang berharga bagi konsumen dalam memilih smartphone yang sesuai dengan kebutuhan dan budget mereka. Dimana konsumen dapat memiliki produk smartphone dengan spesifikasi tinggi dan harga cukup murah seperti produk dari brand Samsung.

Bagi perusahaan, hasil analisis ini dapat berkontribusi dalam membuat keputusan strategis yang lebih akurat terkait pangsa pasar penjualan produk smartphone sebagai upaya meningkatkan daya saing perusahaan di pasar yang dinamis dan kompetitif.

DAFTAR PUSTAKA

- [1] Z. Nanda Aini, dan Mujazi, 2022. "Pengaruh penggunaan smartphone terhadap konsentrasi belajar siswa", *JRTI (Jurnal Riset Tindakan Indonesia)*, vol. 7, no. 3, pp. 467-475. DOI: <https://doi.org/10.29210/30032126000>
- [2] B. Engelbertus Nggalu, dkk., 2022. " Intensitas Penggunaan Smartphone Terhadap Perkembangan Bahasa Anak Usia Dini", *Jurnal Prodi PLS Universitas Nusa Cendana*, vol.1 no. 2. E-ISSN: 2828-5069
- [3] K. Amalia, dkk., 2023. "Analisis Tingkat Penggunaan Handphone pada Siswa Madrasah Aliyah Negeri di Kabupaten Malang-Turen", *Seminar Nasional Pendidikan dan Pembelajaran Ke-6*
- [4] E. Lira, 2023. "Exposing the dark side: phubbing as a detrimental consequence in the digital era", *JPGI (Jurnal Penelitian Guru Indonesia)*, vol. Vol. 8, No. 1, 2023, pp. 121-129 DOI: <https://doi.org/10.29210/0228950005>
- [5] J. Martins, C. Costa, T. Oliveira, R. Gonçalves, and F. Branco, "How smartphone advertising influences consumers' purchase intention," *Journal of Business Research*, vol. 94, no. August 2017, pp. 378–387, 2019.
- [6] E. Bakhshizadeh, H. Aliasghari, R. Noorossana, and R. Ghousi, "Customer Clustering Based on Factors of Customer Lifetime Value with Data Mining Technique (Case Study: Software Industry)," *Int. J. Ind. Eng. Prod. Res.*, vol. 33, no. 1, pp. 1–16, 2022, <https://doi.org/10.22068/ijiepr.33.1.1>
- [7] I. G. A. W. Putri and I. B. G. Dwidasmara, "Application of the K-Means Algorithm to Segmentation of Consumer Interest in Silver Craft 'Kreasi Slaka Bali,'" *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 9, no. 4, p. 541, 2021, doi: [10.24843/jlk.2021.v09.i04.p12](https://doi.org/10.24843/jlk.2021.v09.i04.p12).
- [8] C. Yefta, Yap Rui Qi, " Penerapan K-Means pada Segmentasi Pasar untuk Riset Pemasaran pada Startup Early Stage dengan Menggunakan CRISP-DM," *JURIKOM (Jurnal Riset Komputer)*, Vol. 9 No. 4, 2022, pp. 966-973, DOI. [10.30865/jurikom.v9i4.4486](https://doi.org/10.30865/jurikom.v9i4.4486)
- [9] S. P. Dewi, Nurwati, N., & Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, 3(4), 639–648. <https://doi.org/10.47065/bits.v3i4.1408>
- [10] M. I. Fikri, Sabrila, T. S., & Azhar, Y. (2020). Perbandingan metode naïve bayes dan support vector machine pada analisis sentimen twitter. *SMATIKA Jurnal: STIKI Informatika Jurnal*, 10(02), 71–76. <https://doi.org/10.32664/smatika.v10i02.455>

- [11] S. Kimiagari, S. Keivanpour, and M. Haverila, "Developing a high-performance clustering framework for global market segmentation and strategic profiling," *J. Strateg. Mark.*, vol. 29, no. 2, pp. 93–116, 2021, doi: 10.1080/0965254X.2019.1628099.
- [12] S., Lestari Akmaludin, A., & Badrul, M. (2020). Implementasi Klasifikasi Naive Bayes Untuk Prediksi Kelayakan Pemberian Pinjaman Pada Koperasi Anugerah Bintang Cemerlang. *PROSISKO: Jurnal Pengembangan Riset Dan Observasi Sistem Komputer*, 7(1), 8–16. <https://doi.org/10.30656/prosisko.v7i1.2129>