

IMPLEMENTASI ALGORITMA XGBOOST UNTUK REKOMENDASI UKURAN PAKAIAN DEWASA BERDASARKAN DATA ANTROPOMETRI

Meitiasari Nurlatifah^{1*}, Asriyanik², Didik Indrayana³

^{1,2}Universitas Muhammadiyah Kota Sukabumi, Jl.Syamsudin. SH No.50 S.H, Cikole, Kec. Cikole, Kota Sukabumi, Jawa Barat 43113, Telp. (0266) 218345

Keywords:

XGBoost, ukuran pakaian, antropometri, BMI, machine learning, CRISP-DM.

Correspondent Email:

meitiasarinurlatifah001@ummi.ac.id



Copyright © 2026 The Author(s), Published by Jurnal Informatika dan Teknik Elektro Terapan. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Abstrak. Perkembangan perdagangan fashion secara daring meningkatkan kebutuhan akan sistem yang mampu membantu pengguna menentukan ukuran pakaian secara tepat. Penelitian ini bertujuan mengimplementasikan algoritma Extreme Gradient Boosting (XGBoost) untuk membangun model klasifikasi ukuran pakaian berdasarkan data antropometri. Penelitian menggunakan metode CRISP-DM dengan dataset sebanyak 1.486 data hasil *data cleaning* dari 1.500 responden. Variabel yang digunakan meliputi jenis kelamin, tinggi badan, berat badan, lingkar dada, lingkar pinggang, BMI, dan kategori BMI. Hasil penelitian menunjukkan bahwa model XGBoost memperoleh *accuracy* sebesar 87,25%, dengan rata-rata *precision*, *recall*, dan *F1-score* sebesar 0,88 serta *cross validation accuracy* sebesar 88%. Model berhasil diimplementasikan ke dalam aplikasi berbasis web menggunakan Gradio dan Hugging Face Spaces sehingga mampu memberikan rekomendasi ukuran pakaian secara *real-time*.

Abstract. The rapid growth of online fashion commerce has increased the need for systems that assist users in selecting appropriate clothing sizes. This study implements the Extreme Gradient Boosting (XGBoost) algorithm to develop a clothing size classification model based on anthropometric data. The CRISP-DM methodology was applied using 1,486 valid records obtained after data cleaning. The input variables included gender, height, weight, chest circumference, waist circumference, BMI, and BMI category. The XGBoost model achieved an accuracy of 87.25%, with an average precision, recall, and F1-score of 0.88, while the average cross-validation accuracy reached 88%. The model was successfully deployed as a web-based application using Gradio and Hugging Face Spaces to provide real-time clothing size recommendations.

1. PENDAHULUAN

Perkembangan teknologi informasi dan internet telah mendorong pertumbuhan perdagangan elektronik (*e-commerce*) di berbagai sektor, salah satunya pada industri fashion. Kemudahan berbelanja secara daring membuat masyarakat semakin memilih platform *e-commerce* untuk membeli berbagai kebutuhan, termasuk pakaian. Meskipun memberikan kemudahan dalam proses transaksi, pembelian pakaian secara daring masih menghadapi berbagai kendala, terutama dalam menentukan ukuran pakaian yang sesuai

karena konsumen tidak dapat mencoba pakaian secara langsung sebelum melakukan pembelian [1]. Kondisi tersebut berpotensi menyebabkan ketidakpuasan pelanggan, meningkatnya tingkat pengembalian (*return*) produk, serta menurunkan kepercayaan konsumen terhadap layanan belanja daring [2].

Ukuran pakaian yang digunakan oleh setiap produsen umumnya memiliki standar yang berbeda sehingga konsumen sering mengalami kesulitan dalam memilih ukuran yang tepat. Selain itu, pemilihan ukuran pakaian yang hanya didasarkan pada tinggi badan atau

berat badan belum mampu menggambarkan karakteristik tubuh seseorang secara menyeluruh. Perbedaan proporsi tubuh, seperti lingkaran dada dan lingkaran pinggang, menyebabkan ukuran pakaian yang sesuai bagi setiap individu dapat berbeda meskipun memiliki tinggi atau berat badan yang sama [3]. Oleh karena itu, diperlukan suatu sistem yang mampu memberikan rekomendasi ukuran pakaian berdasarkan karakteristik tubuh pengguna secara lebih objektif.

Salah satu pendekatan yang dapat digunakan adalah pemanfaatan data antropometri. Antropometri merupakan ilmu yang mempelajari dimensi dan proporsi tubuh manusia yang banyak diterapkan dalam bidang ergonomi, kesehatan, maupun industri fashion. Variabel antropometri seperti jenis kelamin, tinggi badan, berat badan, lingkaran dada, lingkaran pinggang, serta Body Mass Index (BMI) dapat digunakan untuk merepresentasikan karakteristik tubuh sehingga mampu meningkatkan ketepatan dalam menentukan ukuran pakaian yang sesuai [4].

Perkembangan *machine learning* membuka peluang untuk membangun sistem rekomendasi ukuran pakaian yang mampu mempelajari hubungan antara data antropometri dengan ukuran pakaian yang digunakan [5]. Berbagai penelitian sebelumnya telah memanfaatkan algoritma seperti K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest, maupun Decision Tree untuk melakukan klasifikasi ukuran pakaian. Namun, sebagian besar penelitian tersebut masih menggunakan variabel antropometri yang terbatas atau belum mengimplementasikan model ke dalam aplikasi yang dapat dimanfaatkan secara langsung oleh pengguna. Selain itu, pemanfaatan algoritma Extreme Gradient Boosting (XGBoost) pada sistem rekomendasi ukuran pakaian berbasis kombinasi fitur antropometri yang lebih lengkap masih relatif terbatas [6].

Extreme Gradient Boosting (XGBoost) merupakan salah satu algoritma *ensemble learning* berbasis *gradient boosting* yang memiliki kemampuan tinggi dalam menyelesaikan permasalahan klasifikasi [7]. Algoritma ini mampu menghasilkan performa prediksi yang baik, mengurangi risiko *overfitting* melalui mekanisme regularisasi, serta bekerja secara efisien pada data berukuran

besar. Dengan karakteristik tersebut, XGBoost dinilai sesuai untuk diterapkan pada sistem rekomendasi ukuran pakaian yang melibatkan berbagai variabel antropometri.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengimplementasikan algoritma Extreme Gradient Boosting (XGBoost) dalam membangun sistem rekomendasi ukuran pakaian berdasarkan data antropometri pengguna. Variabel yang digunakan meliputi jenis kelamin, tinggi badan, berat badan, lingkaran dada, lingkaran pinggang, Body Mass Index (BMI), dan kategori BMI untuk mengklasifikasikan ukuran pakaian ke dalam lima kategori, yaitu S, M, L, XL, dan XXL [8]. Model yang dihasilkan kemudian diimplementasikan ke dalam aplikasi berbasis web menggunakan framework Gradio dan dipublikasikan melalui Hugging Face Spaces sehingga mampu memberikan rekomendasi ukuran pakaian secara real-time.

2. TINJAUAN PUSTAKA

2.1 Sistem Rekomendasi

Sistem rekomendasi merupakan sistem yang dirancang untuk membantu pengguna dalam mengambil keputusan dengan memberikan saran berdasarkan karakteristik atau kebutuhan tertentu [9]. Pada industri fashion, sistem rekomendasi ukuran pakaian bertujuan membantu pengguna menentukan ukuran pakaian yang sesuai berdasarkan karakteristik tubuh sehingga dapat mengurangi kesalahan pemilihan ukuran dan meningkatkan kepuasan pengguna dalam berbelanja secara daring.

2.2 Antropometri

Antropometri merupakan ilmu yang mempelajari dimensi dan proporsi tubuh manusia. Data antropometri banyak dimanfaatkan dalam bidang ergonomi, kesehatan, dan industri fashion untuk menyesuaikan produk dengan karakteristik fisik pengguna [10]. Pada penelitian ini, data antropometri yang digunakan meliputi jenis kelamin, tinggi badan, berat badan, lingkaran dada, lingkaran pinggang, Body Mass Index (BMI), dan kategori BMI sebagai variabel masukan dalam proses klasifikasi ukuran pakaian.

2.3 Machine Learning

Machine learning merupakan cabang kecerdasan buatan (*Artificial Intelligence*) yang memungkinkan sistem mempelajari pola dari data untuk menghasilkan prediksi atau klasifikasi tanpa diprogram secara eksplisit [11]. Pada penelitian ini, machine learning digunakan untuk membangun model klasifikasi yang mampu mempelajari hubungan antara data antropometri dengan ukuran pakaian pengguna.

2.4 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) merupakan algoritma *ensemble learning* berbasis *gradient boosting* yang membangun sejumlah *decision tree* secara bertahap untuk meningkatkan akurasi prediksi [12]. Algoritma ini memiliki keunggulan dalam menangani data tabular, menghasilkan performa klasifikasi yang tinggi, serta dilengkapi mekanisme regularisasi untuk mengurangi risiko *overfitting*. Oleh karena itu, XGBoost dipilih sebagai algoritma utama dalam penelitian ini untuk mengklasifikasikan ukuran pakaian ke dalam lima kategori, yaitu S, M, L, XL, dan XXL.

2.5 CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan metodologi pengembangan data mining yang terdiri atas enam tahapan, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Metodologi ini digunakan sebagai pedoman dalam proses pembangunan model, mulai dari identifikasi permasalahan hingga implementasi model ke dalam aplikasi berbasis web [13].

2.6 Penelitian Terdahulu

Berdasarkan beberapa penelitian terdahulu, pemanfaatan data antropometri dan algoritma machine learning seperti SVM, K-Nearest Neighbor (KNN), Decision Tree, serta Bayesian telah terbukti mampu meningkatkan akurasi klasifikasi ukuran pakaian dan sistem rekomendasi [14]. Penelitian-penelitian tersebut menunjukkan bahwa ukuran tubuh merupakan faktor utama dalam menentukan ukuran pakaian yang sesuai, meskipun masih terdapat keterbatasan, seperti penggunaan fitur antropometri yang belum lengkap dan permasalahan *size and fit* pada e-commerce.

Oleh karena itu, penelitian ini mengembangkan sistem rekomendasi ukuran pakaian menggunakan algoritma XGBoost dengan data antropometri yang lebih komprehensif untuk meningkatkan akurasi prediksi.

3. METODE PENELITIAN

Penelitian ini menggunakan metodologi Cross Industry Standard Process for Data Mining (CRISP-DM) dalam membangun sistem rekomendasi ukuran pakaian menggunakan algoritma XGBoost. Metodologi ini terdiri atas enam tahapan, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment [15]. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Metode CRISP-DM

3.1 Business Understanding

Tahap *Business Understanding* bertujuan untuk mengidentifikasi permasalahan yang terjadi dalam pemilihan ukuran pakaian pada pembelian secara daring. Permasalahan utama yang dihadapi adalah ketidaksesuaian ukuran pakaian karena pengguna tidak dapat mencoba pakaian sebelum melakukan pembelian. Berdasarkan permasalahan tersebut, penelitian ini bertujuan membangun sistem rekomendasi ukuran pakaian berbasis data antropometri menggunakan algoritma XGBoost agar pengguna memperoleh rekomendasi ukuran pakaian yang lebih sesuai.

3.2 Data Understanding

Tahap *Data Understanding* dilakukan untuk memahami karakteristik dataset yang digunakan. Dataset diperoleh melalui penyebaran kuesioner secara daring dan menghasilkan 1.500 data responden. Setelah dilakukan proses *data cleaning*, diperoleh sebanyak 1.486 data yang digunakan dalam penelitian. Variabel yang digunakan meliputi jenis kelamin, tinggi badan, berat badan, lingkar

dada, lingkaran pinggang, Body Mass Index (BMI), kategori BMI, serta ukuran pakaian sebagai variabel target.

3.3 Data Preparation

Tahap *Data Preparation* dilakukan melalui proses pembersihan data, penghapusan data duplikat, pemeriksaan *missing value*, perhitungan BMI sebagai hasil *feature engineering*, pembentukan kategori BMI, serta transformasi data kategorikal menggunakan *label encoding*. Selanjutnya dataset dibagi menjadi data pelatihan dan data pengujian dengan rasio 80:20 menggunakan metode *train-test split* disertai *stratified sampling* untuk mempertahankan proporsi setiap kelas ukuran pakaian.

3.4 Modeling

Tahap Modeling dilakukan menggunakan algoritma Extreme Gradient Boosting (XGBoost). Model dibangun menggunakan variabel jenis kelamin, tinggi badan, berat badan, lingkaran dada, lingkaran pinggang, BMI, dan kategori BMI sebagai fitur masukan untuk mengklasifikasikan ukuran pakaian ke dalam lima kategori, yaitu S, M, L, XL, dan XXL. Model kemudian dilatih menggunakan data pelatihan sebelum dilakukan proses evaluasi.

3.5 Evaluation

Tahap *Evaluation* bertujuan untuk mengukur performa model yang telah dibangun. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, dan *5-fold cross validation*. Selain itu, dilakukan analisis *feature importance* untuk mengetahui kontribusi masing-masing variabel terhadap hasil prediksi ukuran pakaian.

3.6 Deployment

Tahap *Deployment* dilakukan dengan mengimplementasikan model XGBoost ke dalam aplikasi berbasis web menggunakan framework Gradio. Aplikasi kemudian dipublikasikan melalui Hugging Face Spaces, sehingga sistem dapat diakses melalui browser dan memberikan rekomendasi ukuran pakaian secara *real-time* berdasarkan data antropometri yang dimasukkan oleh pengguna.

4. HASIL DAN PEMBAHASAN

Pada bagian ini disajikan hasil implementasi algoritma Extreme Gradient Boosting (XGBoost) dalam membangun sistem rekomendasi ukuran pakaian berdasarkan data antropometri. Pembahasan disusun berdasarkan tahapan CRISP-DM yang meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*.

4.1 Business Understanding

Tahap *Business Understanding* bertujuan mengidentifikasi permasalahan dalam penentuan ukuran pakaian pada pembelian secara daring. Pengguna sering mengalami kesulitan memilih ukuran yang sesuai karena tidak dapat mencoba pakaian sebelum melakukan pembelian. Oleh karena itu, penelitian ini mengembangkan sistem rekomendasi ukuran pakaian berbasis data antropometri menggunakan algoritma XGBoost untuk membantu pengguna memperoleh ukuran pakaian yang lebih sesuai.

4.2 Data Understanding

Tahap *Data Understanding* dilakukan untuk memahami karakteristik dataset yang digunakan dalam penelitian. Dataset diperoleh melalui penyebaran kuesioner daring menggunakan Google Form melalui media WhatsApp dan Instagram. Sebanyak 1.500 data responden berhasil dikumpulkan, kemudian setelah proses *data cleaning* diperoleh 1.486 data yang digunakan dalam proses pembangunan model. Dataset terdiri atas enam variabel, yaitu tinggi badan, berat badan, jenis kelamin, lingkaran dada, lingkaran pinggang, dan ukuran pakaian. Selain itu, penelitian ini membentuk dua fitur baru melalui proses *feature engineering*, yaitu Body Mass Index (BMI) dan kategori BMI.

Tabel 1. Variabel Penelitian

Keterangan	Nilai
Jumlah Responden	1.500
Jumlah Variabel	6

Sebagai gambaran awal, sebagian data hasil pengumpulan kuesioner ditunjukkan pada,

Tabel 2. Contoh Dataset Penelitian

No	TB (cm)	BB (kg)	Jenis Kelamin	Lingkar (cm)	Dada Pin
1	168	85	L	129	118
2	175	79	L	101	117
3	160	50	P	99	97
4	159	49	L	113	106
5	170	59	L	117	108

Karakteristik variabel yang digunakan pada penelitian ditunjukkan pada Tabel 3.

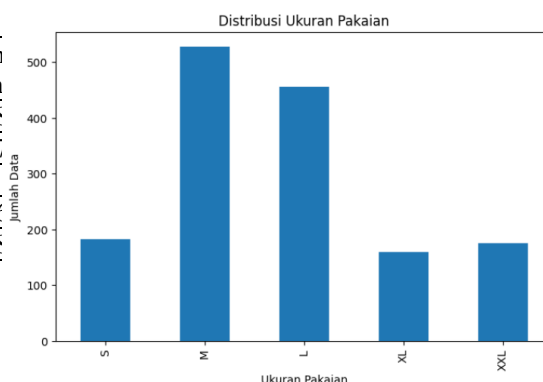
Tabel 3. Karakteristik Variabel Dataset

No	Variabel	Tipe Data	Keterangan
1	Tinggi Badan	Numerik	Tinggi badan responden (cm)
2	Berat Badan	Numerik	Berat badan responden (kg)
3	Jenis Kelamin	Kategorikal	Laki-laki atau perempuan
4	Lingkar Dada	Numerik	Lingkar dada responden (cm)
5	Lingkar Pinggang	Numerik	Lingkar pinggang responden (cm)
6	Ukuran Pakaian	Kategorikal	S, M, L, XL, atau XXL

Hasil analisis statistik deskriptif menunjukkan bahwa data antropometri memiliki variasi yang cukup baik sehingga mampu merepresentasikan karakteristik tubuh responden. Distribusi jenis kelamin juga relatif seimbang, terdiri atas 764 responden perempuan (50,93%) dan 736 responden laki-laki (49,07%). Sementara itu, distribusi ukuran pakaian didominasi oleh kategori M (35,13%) dan L (30,40%), sedangkan kategori S, XL, dan XXL memiliki proporsi yang lebih kecil. Kondisi ini menunjukkan bahwa distribusi kelas belum sepenuhnya seimbang, namun setiap kategori masih memiliki jumlah data yang memadai untuk proses pelatihan model.

Tabel 4. Distribusi Ukuran Pakaian

No	Ukuran Pakaian	Jumlah Data	Persentase (%)
1	M	527	35,13
2	L	456	30,40
3	S	182	12,13
4	XXL	176	11,73
5	XL	159	10,60
Total		1500	100



Gambar 2. Diagram Batang Distribusi Ukuran Pakaian

Selanjutnya dilakukan pemeriksaan kualitas data untuk mengidentifikasi *missing value* dan data duplikat. Hasil pemeriksaan menunjukkan bahwa tidak terdapat *missing value* pada seluruh variabel, sedangkan ditemukan 14 data duplikat yang selanjutnya dihapus pada tahap *Data Preparation*. Dengan demikian, dataset dinilai memiliki kualitas yang baik dan layak digunakan pada proses pemodelan menggunakan algoritma XGBoost.

4.3 Data Preparation

Tahap *Data Preparation* dilakukan untuk mempersiapkan dataset sebelum proses pembentukan model menggunakan algoritma XGBoost. Berdasarkan hasil *Data Understanding*, ditemukan 14 data duplikat yang kemudian dihapus melalui proses *data cleaning*, sehingga jumlah data berkurang dari 1.500 menjadi 1.486 data. Hasil proses *data cleaning* ditunjukkan pada Tabel 5.

Tabel 5. Hasil Data Cleaning

Keterangan	Jumlah Data
Data Awal	1500
Data Duplikat	14
Data Setelah Cleaning	1486

Selanjutnya dilakukan proses *feature engineering* dengan menghitung nilai Body Mass Index (BMI) berdasarkan tinggi badan dan berat badan responden. Nilai BMI kemudian dikelompokkan ke dalam empat kategori, yaitu *Underweight*, *Normal Weight*, *Overweight*, dan *Obesity*. Penambahan fitur BMI dan kategori BMI bertujuan untuk memberikan informasi tambahan mengenai proporsi tubuh sehingga dapat meningkatkan

kemampuan model dalam mengklasifikasikan ukuran pakaian.

Tabel 6. Distribusi Kategori BMI

Kategori BMI	Jumlah Data	Persentase
Normal Weight	1052	70,79%
Overweight	248	16,69%
Underweight	133	8,95%
Obesity	53	3,57%
Total	1486	100,00%

Distribusi kategori BMI menunjukkan bahwa mayoritas responden berada pada kategori *Normal Weight*, sedangkan kategori *Obesity* memiliki jumlah data paling sedikit. Meskipun distribusi belum sepenuhnya seimbang, seluruh kategori masih memiliki representasi data yang memadai untuk proses pelatihan model.

Tahap berikutnya adalah transformasi data kategorikal menggunakan teknik *label encoding*. Variabel jenis kelamin dan kategori BMI dikonversi menjadi data numerik sebagai *feature*, sedangkan ukuran pakaian dikonversi menjadi *label* yang akan diprediksi oleh model XGBoost. Selanjutnya, dataset dibagi menjadi data pelatihan (*training data*) dan data pengujian (*testing data*) menggunakan metode *train-test split* dengan rasio 80:20 serta menerapkan parameter *stratify=y* untuk mempertahankan proporsi setiap kategori ukuran pakaian pada kedua kelompok data.

Tabel 7. Hasil Pembagian Dataset

Jenis Data	Jumlah Data	Persentase
Data Latih (<i>Training Data</i>)	1188	80%
Data Uji (<i>Testing Data</i>)	298	20%
Total	1486	100%

Berdasarkan Tabel 7, sebanyak 1.188 data digunakan sebagai data pelatihan dan 298 data sebagai data pengujian. Penggunaan *stratified sampling* bertujuan mempertahankan proporsi setiap kategori ukuran pakaian sehingga proses pelatihan dan evaluasi model dapat menghasilkan performa yang lebih representatif terhadap distribusi data asli. Dataset yang telah dipersiapkan selanjutnya digunakan pada tahap *Modeling* untuk membangun model klasifikasi menggunakan algoritma XGBoost.

4.4 Modeling

Tahap *Modeling* dilakukan untuk membangun model klasifikasi ukuran pakaian menggunakan algoritma Extreme Gradient Boosting (XGBoost). Dataset hasil *Data Preparation* digunakan sebagai data masukan untuk proses pelatihan model. Variabel *feature* yang digunakan terdiri atas tinggi badan, berat badan, jenis kelamin, lingkar dada, lingkar pinggang, Body Mass Index (BMI), dan kategori BMI, sedangkan variabel target berupa ukuran pakaian yang diklasifikasikan ke dalam lima kategori, yaitu S, M, L, XL, dan XXL.

Model dibangun menggunakan XGBClassifier dengan parameter *objective* multi:softmax untuk klasifikasi multikelas, *num_class* = 5 sesuai jumlah kelas target, dan *random_state* = 42 untuk menjaga konsistensi hasil pelatihan. Parameter model yang digunakan ditunjukkan pada Tabel 8.

Tabel 8. Parameter Model XGBoost

No	Parameter	Nilai
1	objective	multi:softmax
2	num class	5
3	random state	42

Setelah konfigurasi model ditetapkan, proses pelatihan dilakukan menggunakan data latih yang diperoleh dari tahap pembagian dataset. Model dilatih menggunakan 1.188 data pelatihan yang terdiri atas tujuh variabel *feature* dan lima kategori ukuran pakaian sebagai variabel target. Informasi data pelatihan yang digunakan ditunjukkan pada Tabel 9.

Tabel 9. Informasi Data Pelatihan Model

Keterangan	Nilai
Jumlah Data Latih	1.188
Jumlah Fitur	7
Jumlah Kelas Target	5
Algoritma	XGBoost

Model XGBoost dilatih menggunakan 1.188 data latih dengan tujuh variabel *feature* untuk mempelajari hubungan antara karakteristik antropometri dan ukuran pakaian. Model yang telah dilatih selanjutnya digunakan untuk memprediksi data uji, kemudian dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, dan *cross validation* pada tahap berikutnya.

4.5 Evaluation

Tahap *Evaluation* dilakukan untuk mengukur performa model XGBoost menggunakan data uji yang tidak digunakan selama proses pelatihan. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, *5-fold cross validation*, dan *feature importance* untuk mengetahui tingkat ketepatan, konsistensi, serta kontribusi masing-masing variabel terhadap hasil prediksi.

Hasil evaluasi menggunakan metrik *accuracy* menunjukkan bahwa model XGBoost mampu mengklasifikasikan 260 data dengan benar dari total 298 data uji, sehingga diperoleh nilai *accuracy* sebesar 87,25%. Hasil ini menunjukkan bahwa model memiliki tingkat ketepatan yang baik dalam mengklasifikasikan ukuran pakaian berdasarkan karakteristik antropometri pengguna.

Tabel 10. Hasil Accuracy Model XGBoost

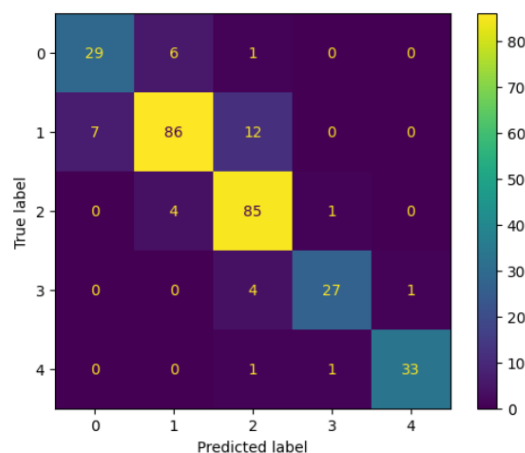
Keterangan	Nilai
Prediksi Benar	260
Total Data Uji	298
Accuracy	87,25%

Selain *accuracy*, evaluasi dilakukan menggunakan *classification report* untuk mengetahui performa model pada setiap kategori ukuran pakaian. Hasil menunjukkan bahwa kategori XXL memiliki nilai *precision* tertinggi sebesar 0,97, sedangkan kategori L memperoleh *recall* tertinggi sebesar 0,94. Nilai *macro average* dan *weighted average* yang berada di atas 0,87 menunjukkan bahwa model memiliki performa yang baik dan relatif seimbang pada seluruh kategori ukuran pakaian.

Tabel 11. Hasil Classification Report Model XGBoost

Ukuran Pakaian	Precision	Recall	F1-Score	Support
S	0,81	0,81	0,81	36
M	0,90	0,82	0,86	105
L	0,83	0,94	0,88	90
XL	0,93	0,84	0,89	32
XXL	0,97	0,94	0,96	35
Macro Average	0,89	0,87	0,88	298
Weighted Average	0,88	0,87	0,87	298

Confusion matrix digunakan untuk mengetahui pola kesalahan klasifikasi pada setiap kategori ukuran pakaian. Hasil menunjukkan bahwa sebagian besar data berada pada diagonal utama matriks, yang menandakan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar. Kesalahan klasifikasi masih terjadi pada beberapa kategori yang memiliki karakteristik antropometri yang berdekatan, terutama antara ukuran M dan L.

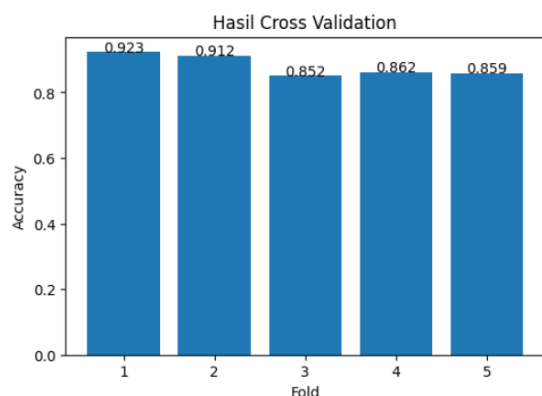


Gambar 3. Confusion Matrix Model XGBoost

Untuk menguji konsistensi model, dilakukan evaluasi menggunakan 5-fold cross validation. Nilai *accuracy* pada setiap *fold* berada pada rentang 85,19%–92,28%, dengan rata-rata sebesar 88,15%. Hasil tersebut menunjukkan bahwa model memiliki kemampuan generalisasi yang baik dan performa yang relatif stabil pada berbagai pembagian data.

Tabel 12. Hasil 5-Fold Cross Validation Model XGBoost

Fold	Accuracy
Fold 1	92,28%
Fold 2	91,25%
Fold 3	85,19%
Fold 4	86,20%
Fold 5	85,86%
Rata-rata	88,15%

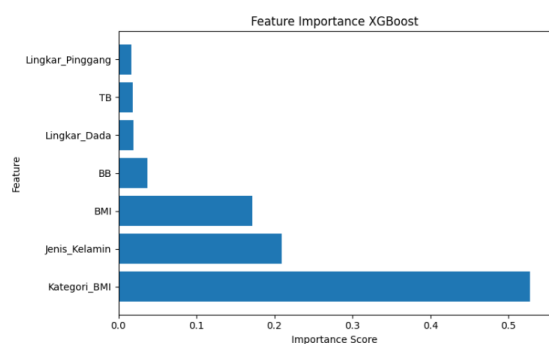


Gambar 3. Diagram Batang Hasil 5-Fold Cross Validation

Analisis *feature importance* dilakukan untuk mengetahui tingkat kontribusi setiap variabel terhadap hasil prediksi. Hasil menunjukkan bahwa Kategori BMI merupakan variabel yang paling berpengaruh dengan nilai *importance* sebesar 0,527372, diikuti oleh Jenis Kelamin sebesar 0,209650 dan BMI sebesar 0,171675. Temuan ini menunjukkan bahwa fitur hasil *feature engineering*, yaitu BMI dan kategori BMI, memberikan kontribusi yang signifikan dalam meningkatkan kemampuan model mengklasifikasikan ukuran pakaian.

Tabel 13. Feature Importance Model XGBoost

Variabel	Importance
Kategori BMI	0,527372
Jenis Kelamin	0,209650
BMI	0,171675
Berat Badan (BB)	0,036870
Lingkar Dada	0,019090
Tinggi Badan (TB)	0,018620
Lingkar Pinggang	0,016724

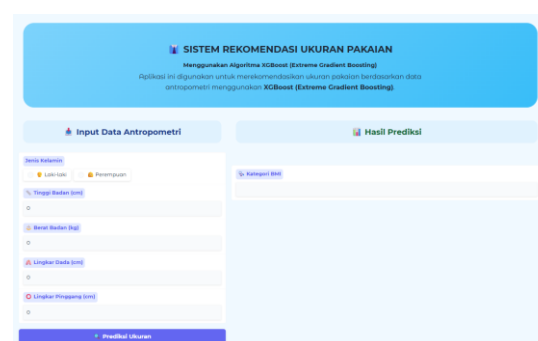


Gambar 4. Diagram Batang Feature Importance

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model XGBoost memperoleh *accuracy* sebesar 87,25%, rata-rata *cross validation accuracy* sebesar 88,15%, serta nilai *precision*, *recall*, dan *F1-score* yang tinggi pada seluruh kategori ukuran pakaian. Hasil tersebut menunjukkan bahwa model memiliki performa yang baik, stabil, dan layak digunakan sebagai dasar pengembangan sistem rekomendasi ukuran pakaian berbasis web.

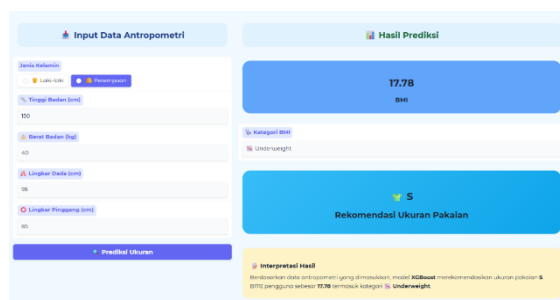
4.6 Deployment

Model XGBoost yang telah melalui proses pelatihan dan evaluasi selanjutnya diimplementasikan ke dalam aplikasi berbasis web menggunakan framework Gradio. Model yang telah disimpan dalam bentuk file kemudian diintegrasikan ke dalam aplikasi untuk menghasilkan rekomendasi ukuran pakaian berdasarkan data antropometri pengguna. Aplikasi dipublikasikan menggunakan platform Hugging Face Spaces sehingga dapat diakses secara daring melalui browser tanpa memerlukan instalasi tambahan.



Gambar 5. Tampilan Halaman Utama Sistem

Berdasarkan Gambar 5, pengguna dapat memasukkan data berupa jenis kelamin, tinggi badan, berat badan, lingkar dada, dan lingkar pinggang. Sistem kemudian menghitung nilai Body Mass Index (BMI) beserta kategori BMI secara otomatis sebelum melakukan proses prediksi menggunakan model XGBoost.



Gambar 6. Tampilan Hasil Prediksi Sistem

Berdasarkan Gambar 6, sistem berhasil menampilkan hasil prediksi secara *real-time* berupa rekomendasi ukuran pakaian, nilai BMI, kategori BMI, dan interpretasi hasil berdasarkan data antropometri yang dimasukkan pengguna. Aplikasi telah berhasil dipublikasikan melalui Hugging Face Spaces dan dapat diakses melalui alamat <https://meitiassss-rekomendasi-size.hf.space/>. Implementasi ini menunjukkan bahwa model XGBoost tidak hanya memiliki performa yang baik pada tahap evaluasi, tetapi juga berhasil diterapkan menjadi aplikasi berbasis web yang dapat digunakan secara langsung untuk membantu pengguna menentukan ukuran pakaian yang sesuai.

5. KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma Extreme Gradient Boosting (XGBoost) untuk membangun sistem rekomendasi ukuran pakaian berdasarkan data antropometri. Model yang dibangun menggunakan variabel tinggi badan, berat badan, jenis kelamin, lingkar dada, lingkar pinggang, BMI, dan kategori BMI memperoleh *accuracy* sebesar 87,25%, dengan rata-rata *precision*, *recall*, dan *F1-score* sebesar 0,88, serta rata-rata *cross validation accuracy* sebesar 88,15%. Model juga berhasil diimplementasikan ke dalam aplikasi berbasis web menggunakan Gradio dan Hugging Face Spaces sehingga mampu memberikan rekomendasi ukuran pakaian secara *real-time* berdasarkan data antropometri pengguna.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Allah SWT, dosen pembimbing, keluarga, serta semua pihak yang telah memberikan dukungan,

bimbingan, dan motivasi sehingga penelitian ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

- [1] R. Alhassawi, S. Gill and d. , "Evaluating machine learning models for clothing size prediction using anthropometric measurements from 3D body scanning," 2025.
- [2] B. B. Mumtazah and S. D. Sancoko, "Rekomendasi Ukuran Baju Dewasa Menggunakan Algoritma K-Nearest Neighbor dan Support Vector Machine," *Institut Riset dan Publikasi Indonesia(IRPI)*, 2024.
- [3] X. Lawrenza, H. Fitriyah and d. , "Rancang Bangun Sistem Klasifikasi Ukuran Baju berdasarkan Ukuran Tubuh dengan Metode K-Nearest Neighbor berbasis Arduino," *Jurnal pengembangan teknologi informasi dan ilmu komputer*, 2022.
- [4] T. N. Laily and N. Pratiwi, "Klasifikasi Ukuran Baju Berdasarkan Pengukuran Tubuh Menggunakan MediaPipe dan Support Vector Machine," *Metik Jurnal*, 2025.
- [5] M. N. P. Sari and d. , "Implementation of Anthropometry on the Design of Study Tables and Chairs in Pekanbaru 17 Public Elementary School Students Against the Prevention of Musculoskeletal Disorders (MSD's)," 2022.
- [6] R. Maharesi, "Implementasi Algoritma Decision Tree Cart Untuk Merekomendasikan Ukuran Baju," 2002.
- [7] D. B. saputra, V. Atina and F. E. Nastiti, "PENERAPAN MODEL CRISP-DM PADA PREDIKSI NASABAH KREDIT MENGGUNAKAN ALGORITMA RANDOM FOREST," 2024.
- [8] s. suhada, S. Bahri and d. , "Sistem Rekomendasi Produk Menggunakan Metode User-BasedCollaborative Filtering Pada Digital Marketing," *J-INTCH (Journal of Information and Technology)*J-INTCH (Journal of Information and Technology)J-INTCH (Journal of Information and Technology)J-INTCH (Journal of Information and Technology), 2022.
- [9] A. Rachmaniar, "SISTEM REKOMENDASI PRODUK E-COMMERCE MENGGUNAKAN COLLABORATIVE FILTERING DAN CONTENT-BASED FILTERING," 2025.
- [10] I. W. G. Suarjana, "PERANCANGAN FASILITAS KERJA ERGONOMI MENGGUNAKAN DATA ANTROPOMETRI UNTUK

MENGURANGI BEBAN FISILOGIS,”
2022.

- [11] A. Sutaryani, S. Sunarno and D. ,
“PERBANDINGAN PERFORMA MODEL
MACHINE LEARNING DALAM
PREDIKSI SUHU DI SEMARANG,” *Jurnal
Informatika dan teknik elektro terapan*, 2024.
- [12] F. P. ari and R. B. , “Implementasi Extreme
Gradient Boosting (XGBoost) untuk
Klasifikasi Penggunaan Ulang Asisten AI,”
2025.
- [13] D. Ruswanti, D. Susilo and d. , “Implementasi
CRISP-DM pada Data Mining untuk
Melakukan Prediksi Pendapatan dengan
Algoritma C.45,” 2024.
- [14] D. M. Silalahi and M. F. , “Algoritma Decision
Tree dalam Menentukan Rekomendasi
Pakaian Berdasarkan Fitur Fisik dan
Preferensi Gaya,” 2025.
- [15] D. Wilandini and P. , “PENERAPAN
ALGORITMA NAÏVE BAYES DALAM
MENGKLASIFIKASIKAN MEDIA SOSIAL
UNTUK MENGAMATI TREND
KULINER,” *Jurnal Teknologi Terpadu*, 2022.